



```
Db 61 IMMKEKEHTNLMSTLKKREEKQKQALKNLNEVQEHLEEBERLCRESLADSWGCRSCL 120
Qy 150 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKDAQLTQM 209
Db 121 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKDAQLTQM 180
Qy 210 EDVFSQLTVDVNSLNRSGFQKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 269
Db 181 EDVFSQLTVDVNSLNRSGFQKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 240
Qy 270 DLEQCDWIPNFFQLCNFSVSVIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 329
Db 241 DLEQCDWIPNFFQLCNFSVSVIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 300
Qy 330 DRGLCGELDONLSRCKFHEKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 389
Db 301 DRGLCGELDONLSRCKFHEKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 360
Qy 390 LQMKRKHLEDATYLVKEMRGQFGWVSELANOAPETETIFNSIQVVPRIHEGNSKQDETM 449
Db 361 LQMKRKHLEDATYLVKEMRGQFGWVSELANOAPETETIFNSIQVVPRIHEGNSKQDETM 420
Qy 450 MTDLSILPSSNFTLKIPLAESAESNFIYGYVAKALQHEKHEFTW 495
Db 421 MTDLSILPSSNFTLKIPLAESAESNFIYGYVAKALQHEKHEFTW 466

RESULT 2
Q95KNI PRELIMINARY; PRT; 465 AA.
ID AC Q95KNI
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE RETINAL CLUSTERIN-LIKE PROTEIN CLULB SPICE VARIANT.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF241221; AAK49030.1; -.
SQ SEQUENCE 465 AA; 54402 MW; BBDE1AC512D5D33F CRC64;

Query Match 69.6%; Score 1832.5; DB 6; Length 465;
Best Local Similarity 73.2%; Pred. No. 1.1e-117;
Matches 341; Conservative 52; Mismatches 72; Indels 1; Gaps 1;

Qy 30 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALGTQKQK 89
Db 1 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALGTQKQK 60
Qy 90 IMMKEKEHTNLMSTLKKREEKQKQALKNLNEVQEHLEEBERLCRESLADSWGCRSCL 149
Db 61 IMMKEKEHTNLMSTLKKREEKQKQALKNLNEVQEHLEEBERLCRESLADSWGCRSCL 120
Qy 150 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKDAQLTQM 209
Db 121 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKDAQLTQM 180
Qy 210 EDVFSQLTVDVNSLNRSGFQKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 269
Db 181 ENVFNQLTVDVNLNRSLNVRFMQOEFDQTFQSFYMSDITLQPNFLPALSKEPRKKA 240
Qy 270 DLEQCDWIPNFFQLCNFSVSVIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 329
```

```
Db 241 DPVQSWDIPSFQFLYFNFSLSIYHSITTTITKTLNAIEDLPKQDNDNSHGSLSKTLPVQ 300
Qy 330 DRGLCGELDONLSRCKFHEKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOI 389
Db 301 HRGYPGEQGNLSECFQFHARQKQCDYLWEDCPDVELHTKVDLEALYNISHQQAQV 360
Qy 390 LQMKRKHLEDATYLVKEMRGQFGWVSELANOAPETETIFNSIQVVPRIHEGNSKQDETM 449
Db 361 LQMKRKHLEDATYLVKEMRGQFGWVSELANOAPETETIFNSIQVVPRIHEGNSKQDETM 420
Qy 450 MTDLSILPSSNFTLKIPLAESAESNFIYGYVAKALQHEKHEFTW 495
Db 421 I-DLSILSPNFTLKIPLAESAESNFIYGYVAKALQHEKHEFTW 465

RESULT 3
Q9NIT8 PRELIMINARY; PRT; 338 AA.
ID AC Q9NIT8
DT 01-OCT-2000 (TremBLrel. 15, Created)
DT 01-OCT-2000 (TremBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TremBLrel. 17, Last annotation update)
DE RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.
GN CLUL1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF147784; AAF36799.1; -.
DR InterPro; IPR000753; Clusterin.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
SQ SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;

Query Match 46.6%; Score 1226.5; DB 6; Length 338;
Best Local Similarity 69.8%; Pred. No. 2.2e-76;
Matches 231; Conservative 42; Mismatches 57; Indels 1; Gaps 1;

Qy 165 SSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKDAQLTQMEDVFSQLTVDVNSLF 224
Db 9 TSGTSTVQGFNIIYQFLPFDEDEKDLPVGEKFIEDAQAQVNIENFQLTVDVREL 68
Qy 225 NRSFNVRMQOEFDQTFQSFISDTDLTEPYFFPAFKEPMTKADLCQCDWIPNFFQLF 284
Db 69 NRSFNVRMQOEFDQTFQSFYMSDITLQPNFLPALSKEPRKADPVQSWDIPSFQFL 128
Qy 285 CNFSYIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGODRGLCGELDONLSRC 344
Db 129 YNFSLSIYHSITTTITKTLNAIEDLPKQDNDNSHGSLSKTLPVQHRGYPGEQGNLSEC 188
Qy 345 RFHFHCKQKQAHLSQEDCPDVPALHTELDEAIRLVNVSNOQYGOIQLMKRKHLEDATYLV 404
Db 189 FQFHARQKQCDYLWEDCPDVELHTKVDLEALYNISHQQAQVQLQMTQHLEDITVLM 248
Qy 405 EKMKGQFGWVSELANOAPETETIFNSIQVVPRIHEGNSKQDETMKTDLSILPSSNFTLK 464
Db 249 EKMKEEGWVADLANQAPGAENIFDSTKMPNIIHEGNSKQDETM I-DLSILSPNFTLK 307
Qy 465 IPLESAESNFIYGYVAKALQHEKHEFTW 495
Db 308 IPLESAESNFIYGYVAKALQHEKHEFTW 338

RESULT 4
```

```

Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OLIVER; TISSUE=TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC EMBL; L46797; AAA80313.1; .
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART; SM00035; CLb; 1.
DR SMART; SM00030; CLb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
FT SEQUENCE 449 AA; 52154 MW; 705A605045157C0 CRC64;

Query Match 16.1%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 3.2e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps

Qy 34 LLVFIIVCLLWKSHCAPTWKDKATISENLKSFSEVGEIDAEVVKALTGIKQKIMME 93
Db 4 LLLVLGLLLTLENGQ----VLGDRAVSDELQENSTQGSRYINKEINKNALGVYKIKNLE 60
Qy 94 RKEKHTNLMSTLKKREERQOEALKLLNEVQHELEERLCRESLADSGECRSGLENNC 153
Db 61 QTNEERKSLTGLEAAKKKEGALNDTKDSEMKLKESQGVCMETWALWEECPCLQKTC 120
Qy 154 MRYIT-TCQPSWSVYVKNIEFRFRKYYQFLFPFHEDNEKDLPISEKLIKIEKDAQLTQMEDV 212
Db 121 MKFYARVCRSGSLGVHQLEEFNLQSSPFYFINGDR-----IDSLLENDRQQTHVLVD 174
Qy 213 FSQLTVDVNSLFRSFNVFMQOQEDFTQFSHFISDTELTPYEPFAPSKPEPTKADL- 271
Db 175 -----MQDSFDPRASSI-----MDELFDQRFTT-REPQDTYYISPPFS-SPHRRSLL 218
Qy 272 -----ECWDIP-----NFFQLFCNFSVIYESVSITIKMLKAIEDLPKQD--KAP 316
Db 219 FNPKSRFARIMHFPYRHLNFMDFMQPFDMTHQAQQAAMNHLHLRLPDQLPMTFESEGD 278
Qy 317 DHGGLISKMLPGDGRGLCGELDNLSRCFKFEKCKQCOAHLSEDC----PDVPAHLTEL 372
Db 279 NH-----DRTVCKEIRHINSTGCLUMKDKQCEKQCEILSVDCSTNNPFSOMQLROEL 327
Qy 373 DEAIRLVNSNOQYGOIGLQMKRKHLEDYALVLEKVRMGQFGWVSELANOAPETEIFNSIQ 432
Db 328 NNSLQLAEKFTKLYDELLQSYQEKMLNTSSLKQLNQEFSWVSQLANLTQGEDQYYIQLVT 387
Qy 433 VVPIRIHEGNIKSKODETMMDTSLTSLPSSNFTLKPILEESAESSNFTGYGVAKALQHFKE 490
Db 388 TVSS-INSNSEVSPGUTRVVVVKKLFQSYPTVTPVPEVSRNNPKFMETVAEKAQOEYR 444

```









SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match  
Best Local Similarity 5.7%; Score 149.5; DB 6; Length 840;  
Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;  
QY 57 TAISENLKSFSEVGEIDAEEVKALTGKQKIMMERKEKHTNLTSLKCKRE-EKQE 115  
DB 426 TEMQKNKSVQY--LEMDTKLSKKEEVKFLQKREQEKVTASADLLKREKETQE 483  
QY 116 ALKLLNEVQE---HLEEEELRESLADSGWGCRCLENNCMRIYTCQPSWSVKNKI 171  
DB 484 FLSQEEFPKRDKNLEERQKL-----KSLR 509  
QY 172 ERFRKIYQFLPFPHEDNEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNSLFRSF--- 228  
DB 510 EKLTVKNLQFMSENERAKNIKQO-----QINEVKNENKLLKHVARSEEQNVPK 562  
QY 229 NVFQMOQEFDTQSFHISDTDLTEPYFFPAFS---KEPMTRKADLE----- 272  
DB 563 SETAQLKEOLEEVNKMSTDTKTHSNLLDCSPCEESINPADIERSQOLASKMHSLL 622  
QY 273 -----QCWDIPNFQLFNFSYISYSEVITMKAL-----EDLPQDKAP 316  
DB 623 ALMVGILLKQDITNSDAE--HFKESE--EKVSDIMLQRLKSLHLKKNLDRKLLKHKDRIT 678  
QY 317 DHGGLISKLMPGQDRGL-----CGE----- 336  
DB 679 TFRDLIAKAFQDHAIKVTDCSDSEAKSIRDVPTFLGAKLDKYHSLNEBELFLIYKLG 738  
QY 337 LDQNLRCFKFHEKQKQKQAHLSQEDCPDVPALHTLDEAIRLVNSNQYQGIQMTKR 395  
DB 739 LLESKESHCHNRLTEENDKYRHGLSLIKVTYSVEIEECADQRLAISHSQIAH-LEKRNK 757  
QY 396 HLEDTAYLVKMRGQFGWVSELANOAPETIEFN---SIOVPRHIEGNSKOD 446  
DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHFKSMTMPFAVENRNDLD 840

RESULT 13

ID Q28021 PRELIMINARY; PRT; 1388 AA.  
AC Q28021;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE RHO-ASSOCIATED KINASE.  
OS Bos taurus (Bovine).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;  
OC Bovidae; Bovinae; Bos.  
OX NCBI\_TaxID=9913;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=96208507; PubMed=8641286;  
RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,  
RA Nakano T., Okawa K., Iwamatsu A., Kabuchi K.;  
RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative  
RT target for small GTP binding protein Rho.";  
RL EMBO J. 15:2208-2216(1996).  
RC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.  
DR EMBL; U36909; AAC48567.1; -;  
DR HSSP; O63450; 1A06.  
DR InterPro; IPR002219; DAG\_PE-bind.  
DR InterPro; IPR000719; Euk\_pkinase.  
DR InterPro; IPR001849; PH.  
DR InterPro; IPR000961; Pkinase.C.  
DR InterPro; IPR000861; REM\_repeat.  
DR InterPro; IPR002290; Ser\_thr\_pkinase.  
DR Pfam; PF00130; DAG\_PE-bind; 1.  
DR Pfam; PF02185; HRI; 1.  
DR Pfam; PF00169; PH; 1.

DR Pfam; PF00069; pkinase; 1.  
DR SMART; SM00109; CI; 1.  
DR SMART; SM00074; HRI; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00220; S\_TKC; 1.  
DR SMART; SM00133; S\_TK\_X; 1.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE; PS00003; PH\_DOMAIN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 1388 AA; 160799 MW; CA83CE7D3860465D CRC64;

Query Match  
Best Local Similarity 5.7%; Score 149; DB 6; Length 1388;  
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 54 KDKTAISENLKSFSEVGEIDAEEVKALTGKQKMI-----MMERKEKEHTNLS 104  
DB 780 KQDVLNEDVRNLT-----LKIEQETQKRCGLTQNDLKMOTQOVNTLKMSEKQLKQENHLL 835  
QY 105 TLK-----KCREKQKQKALKLLNEVQEHLEEE-----RLCR 135  
DB 836 EMKMSLEKQNAELKREKQADGQMKELQOLEAEQFSTLYKTQVRELKEBECEETKCLK 895  
QY 136 E-----SLADSGWGCRCLENNCMRIYTCQPSWSVKNKIERFRKIYQFLPFPHEDN 189  
DB 896 ELQOKKQELQDE-----RDSLAQOLEITLTADSEQLARSIAEQYSDLEKE 942  
QY 190 E--KDLPISE-----KLEKDAQLTQMEDVFSQLTVDVNSLFRSFNVFQMOQEFQ 240  
DB 943 KIMKELEIKEMARHKOELTEKDATIASLEETNRLTSDVANLANEKEELNKLKEAQO 1002  
QY 241 TQSFHISDTLTPYFFPAFSKEPMTRKADLEQCDIPNFQFQFCNFSVSVSEVIT 300  
DB 1003 LSR---LKDEISAAAKAQAQEFKQLLTERTLK-----TQAVN 1036  
QY 301 KMLKATE-----DLPKQDKAPDHGGLISKMLPGQDRGLCGEL-----DQNLSCRF 345  
DB 1037 KLAIEINRKEPVKRGNDTVDVRKEK-----ENRKLHMLKSEREKLITQOMI 1082  
QY 346 KFEHEKQKQKQAHLSQEDCPDVPALHTLDEAIRLVNSNQYQGIQMTKRKHE----- 398  
DB 1083 KYQKELNEMQAIAEE---SQRIEQLQMTLDSKSDIEQLRSQALHIGLDSSSISG 1138  
QY 399 -----DTAYLVKMRG-----QFGWVSEL----- 417  
DB 1139 PGDTEADGFPESRLEGWLSLVPNRNTRKFGWVKYVIVSVSSKILFYDSEQDKQSNPYM 1198  
QY 418 -----ANQAPETEIFNSIQVVPRI-----HEGNISKQDETMMTDLSILPSSNFT 462  
DB 1199 VLDDIKLFHVRPVTQTDVIRADAKIPIRQILYANEGESKEQE----- 1243  
QY 463 LKIPLESAESSNFI---GYVAKALQHF 488  
DB 1244 --FPVEPVGKSNYICHKGHEFIPLYHF 1270

RESULT 14

ID Q18082 PRELIMINARY; PRT; 1091 AA.  
AC Q18082;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 127.0 KDA PROTEIN.  
GN C18C4.5.  
OS Caenorhabditis elegans.  
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;  
OC Rhabditidae; Peloderinae; Caenorhabditis.  
OX NCBI\_TaxID=6239;



```

Best Local Similarity 19.5%; Pred. No. 0.27;
Matches 104; Conservative 96; Mismatches 167; Indels 166; Gaps 23;

QY      62 NLSFSFSGEIDADE--VKKALTGIKOMTKMWERKEKHEHTNLMSTLKK---CREEKOEAA 116
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2378 NMREQKIISLLSKEEAIOVAIAELRQOH--DKYEKENLLSQEEENIVLEEENKKA 2434
QY      117 LKLLNEVOEHLE--EERLCRESLASDWGCRSCLENNCMRI---YTTCQPSSSV--- 167
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2435 VDTNQMLMETLKTIKENIQOAKQSFKVMSLSQNDRRDIQVDGYOOLEERHLSILEX 2494
QY      168 -----KNKIERFFRKIYQFLPFPHEDNKDLPISEKLIE-----DAQL 206
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2495 DQLIQEAAAENNKLKEIRGLSRHMDDLSENAK---LDRELQYREDLNQVITIKDSQQ 2551
QY      207 TQ-----MEDVFSOL-----TVDVNSLFNRNVFRMQOEFDQTQSFHSID 249
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2552 KQLEVOLOONKELENKYAKLEKLEKESEANEDLRSRFNALQEEKODLSKETESLKVSI 2611
QY      250 TDLT-----EPYPPEPA-FSKPEPTWKADLECQWDIPNFQOLF 284
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2612 SOLTQRTVALQBSGTGLGHYAQLKVKEEVHRHLSALFSSSQKRIAEEEE---ELV 2663
QY      285 CNFSVIYESVETITKMKLAIED-LPKODKAPDHGGLSK----- 324
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2664 C-----VQKEAAKVGEIEDKLKELKHLHHDAQIMRNETETAERVAELARDLYE 2714
QY      325 -----MLPGDRGLCGELDQNLSRCFKFEHKCOKQAHLSDCDPDVPALHTELDEAIRL 378
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2715 MEOKLLMVTKENKGTLTAQI-QSFGRSK-----SSLQNSRDHANE-----ELDELKRR 2760
QY      379 VNVSNOQYQGIIOWTRKHLEDATYILVEKMRGOGFWGYSLANQAPETEIIFNSIOVVPRTH 438
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2761 YDASLUKELAQLKEQGLLNRRDALL-----SETAFSMNSTE-ENSLSHLEKLN 2807
QY      439 EGNISKODETMTDLSILPSSNFTLKIPLEESAESSNFIYGYYAKALQHFKKH 491
Db      : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Dbb     2808 QQLLSKDEOLL-----HLSQOLESYNQOVGSFKAMA-SLQNERDH 2847

```

Search completed: July 2, 2002, 11:57:10  
Job time: 408 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:49:54 ; Search time 53.01 Seconds  
(without alignments)  
1556.661 Million cell updates/sec

Title: US-09-722-544A-4  
Perfect score: 2529  
Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHKEHFKTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SP-archaea.\*
- 2: SP-bacteria.\*
- 3: SP-fungi.\*
- 4: SP-human.\*
- 5: SP-invertebrate.\*
- 6: SP-mammal.\*
- 7: SP-mhc.\*
- 8: SP-organelle.\*
- 9: SP-phage.\*
- 10: SP-plant.\*
- 11: SP-rodent.\*
- 12: SP-virus.\*
- 13: SP-vertebrate.\*
- 14: SP-unclassified.\*
- 15: SP-virus.\*
- 16: SP-bacteriap.\*
- 17: SP-archaeap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2464	97.4	466	4	Q15846
2	1836.5	72.6	465	6	Q95KN1
3	1230.5	48.7	338	6	Q9N1T8
4	423	16.7	449	6	Q29482
5	401.5	15.9	448	13	Q9YGP0
6	321	12.7	372	11	Q9JK98
7	266.5	10.5	295	11	Q35510
8	204	8.1	218	11	Q9ERD1
9	160	6.3	1738	5	Q76329
10	160	6.3	3616	13	Q9W6V0
11	151.5	6.0	1091	5	Q18082
12	149.5	5.9	840	6	Q95JRO
13	146	5.8	1156	16	Q66878
14	145	5.7	1388	6	Q28021
15	144	5.7	1305	10	Q9FJ35
16	144	5.7	3259	4	Q14789

17 143 5.7 1931 5 09NCF9  
18 143 5.7 1931 5 09VKH9  
19 142.5 5.6 333 6 095JY2  
20 141 5.6 1388 4 075116  
21 141 5.6 1388 4 090QN5  
22 140.5 5.6 2166 16 051465  
23 140 5.5 1379 11 062868  
24 139 5.5 1330 6 097961  
25 139 5.5 1530 4 043241  
26 139 5.5 1955 5 061308  
27 139 5.5 5458 5 090459  
28 138.5 5.5 1391 11 0922J3  
29 137.5 5.4 720 4 09H6Q7  
30 137 5.4 1084 16 083423  
31 137 5.4 1300 4 013999  
32 136.5 5.4 709 5 09GRG1  
33 136 5.4 1132 4 075065  
34 135.5 5.4 470 10 038843  
35 135.5 5.4 684 3 007238  
36 134.5 5.3 882 17 096YR5  
37 134.5 5.3 1708 5 09U0S6  
38 134.5 5.3 1723 2 09JMX8  
39 134.5 5.3 1958 5 096062  
40 134 5.3 1356 4 014707  
41 133.5 5.3 1956 5 020641  
42 132 5.2 1033 4 09UFEL  
43 132 5.2 1133 5 021022  
44 132 5.2 1270 4 096JN2  
45 132 5.2 1388 11 P70336

#### ALIGNMENTS

RESULT 1  
Q15846  
•ID Q15846 PRELIMINARY; PRT; 466 AA.  
AC Q15846;  
DT 01-NOV-1996 (TREMELrel. 01, Created)  
DT 01-NOV-1996 (TREMELrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMELrel. 19, Last annotation update)  
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=NEURORNETINA;  
RA Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,  
RA Matsubara K.;  
RT "Expression profile of active genes in human retina.";  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D63813; BAA09882.1; -;  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 3.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
KW Signal; Hypothetical protein.  
FT SIGNAL 1 20  
FT CHAIN 21 466 ROD PHOTORECEPTOR PROTEIN.  
SQ SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;

Query Match 97.4%; Score 2464; DB 4; Length 466;  
Best Local Similarity 100.0%; Pred. No. 3.3e-161;  
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 12 MKPPLLVFVCLLWLDKSHCAPTKDKTAISENLKSFSEVGIDEADEEVKALTGIKQMK 71  
|||||  
Db 1 MKPPLLVFVCLLWLDKSHCAPTKDKTAISENLKSFSEVGIDEADEEVKALTGIKQMK 60  
|||||  
QY 72 IMMERKEHTNLMSTLKKCKREEKQEAALKLLNEVOEHLLEERLCRESLADSGECRSL 131

Db	61	IMMERKEHTNLMSTLKKCREKQBALKLLNEVQHLEEBERLCRESLADSWGECRSL	120
QY	132	ENNCMRIYTTCPSSWSVKNKIERFRKTYQFLFPFHEDNEKDLPISEKLTIEDAQLTQM	191
Db	121	ENNCMRIYTTCPSSWSVKNKIERFRKTYQFLFPFHEDNEKDLPISEKLTIEDAQLTQM	180
QY	192	EDVFSQLTVDVNSLFRNSNFRQMQOEFDQTFQSHFISDTDLTPPYFFPAFSKEPMTKA	251
Db	181	EDVFSQLTVDVNSLFRNSNFRQMQOEFDQTFQSHFISDTDLTPPYFFPAFSKEPMTKA	240
QY	252	DLEQCHDIPNFFOLFCNFSVSYESVETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ	311
Db	241	DLEQCHDIPNFFOLFCNFSVSYESVETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ	300
QY	312	DRGLGELDQNLRSRCFKFHEKQKQCAHLSDECDDVPALHTELDEAIRLVNYSNOOYGOI	371
Db	301	DRGLGELDQNLRSRCFKFHEKQKQCAHLSDECDDVPALHTELDEAIRLVNYSNOOYGOI	360
QY	372	LQMKRKHLEDATYLVEKMRQGFQGWSELANQAPETIIFNSIQVVPRIHEGNSIKODETM	431
Db	361	LQMKRKHLEDATYLVEKMRQGFQGWSELANQAPETIIFNSIQVVPRIHEGNSIKODETM	420
QY	432	MTDLSILPSSNFTLKTIPLSESAESSNFYGVYVAKALQHFKEHFKTW	477
Db	421	MTDLSILPSSNFTLKTIPLSESAESSNFYGVYVAKALQHFKEHFKTW	466
RESULT	2		
Q95KN1	ID	Q95KN1 PRELIMINARY;	PRT; 465 AA.
AC	Q95KN1		
DT	01-DEC-2001	(TEMBLrel. 19, Created)	
DT	01-DEC-2001	(TEMBLrel. 19, Last sequence update)	
DT	01-DEC-2001	(TEMBLrel. 19, Last annotation update)	
DE	RETINAL CLUSTERIN-LIKE PROTEIN CLU1LB SPICE VARIANT.		
OS	Canis familiaris (Dog)		
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
OX	NCBI_TaxID=9615;		
ON	[1]		
RP	SEQUENCE FROM N.A.		
RC	TISSUE=RETINA;		
RA	MEDLINE=20156379; PubMed=10675623;		
RX	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;		
RT	"Molecular cloning, characterization and expression of a novel retinal		
RL	clusterin-like protein cDNA.;"		
RL	Gene 243:151-160(2000).		
DR	ENBL; AF241221; AAK49030.1; -.		
SO	SEQUENCE 465 AA; 54402 MW; BBDE1AC512D5D33F CRC64;		

Db	241	DPVQWDIPSFOLFNFSLSIYHSITSTITTKLNAIEDLPQDNDNSHGSLSSKTLPVQ	300
Qy	312	DRGLCGELQNLSCRCFKFHEKCKQAHLSDECDPVPALHTELDEAIRLVNYSNQYGOI	371
Db	301	HRGPYGEFGNLSECFQFARCKQCDYLWEDCDPDPPELHTKVDLEALVNIHQYQAV	360
Qy	372	LQMTKRHLEDATYLVEKMRGCFQFWSELANQAPETELIIFNSIQVVPRIHEGNIKSQDETM	431
Db	361	LQMTQHHLEDTTYLMEKMRFEFGWADLANQAPGAENIFDSTKMPNTHIEGNSKQDETM	420
Qy	432	MTDLSTILSPSNFTLKLPLEBSAESSNFIGVVAKALQHFKEHFKTW	477
Db	421	I-DLSILSSPNFTLKLPLEBSAETSNFISYMLEKAVQHFKEHFKTW	465
RESULT	3:		
Q9N1T8		PRELIMINARY;	PRT; 338 AA.
AC	Q9N1T8;		
DT	01-OCT-2000 (TrEMBLrel. 15, Created)		
DT	01-OCT-2000 (TrEMBLrel. 15, Last sequence update)		
DE	01-JUN-2001 (TrEMBLrel. 17, Last annotation update)		
DT	RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.		
GN	CLU1.		
OS	Canis familiaris (Dog).		
OC	Eukaryota; Metazoa; Chordata; Cranata; Vertebrata; Euteleostomi;		
OC	Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.		
NCBI_TaxID=	9615;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RX	MEDLINE=20156379; PubMed=10675623;		
RT	Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Agulre G.D.;		
RT	"Molecular cloning, characterization and expression of a novel retina		
RT	clusterin-like protein cDNA.;"		
RL	Gene 243:151-160(2000).		
DR	EMBL: AF147784; AAF36799.1; -		
DR	InterPro: IPR000753; Clusterin.		
DR	SMART: SM00035; Clu; 1.		
DR	SMART: SM00030; Clu; 1.		
SQ	SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;		

Qy 2521



```

Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (Tremblrel. 01, Created)
DT 01-NOV-1996 (Tremblrel. 01, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=OLIVER; TISSUE=TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
Clusterin."
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; LA6797; AAA80313.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLa; 1.
DR SMART; SM00030; CLb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
SQ SEQUENCE 449 AA; 52154 MW; 705A605045157C0 CRC64;

Query Match 16.7%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 2.8e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps

QY 16 LLVPIVCLLWLKSHCAPTWKDKTAISENLKPSFEVEIDAEVEVKALTGIKQKIMME 75
Db 4 LLLVLGULLTLENGQ---VLGDRAVSDRELQENSTQGSNYINKENKALKGVYKIKLIE 60
QY 76 RKEKHTNLMLSTLKKREEKQEAALKLLNEVOEHLEERELCRLESADSGEGRSCLENNC 135
Db 61 QTNEERKSLLTLEAAKKKKEGALNDTKDEMKLKESQGVCHETWALWEECKPCLKQTC 120
QY 136 MRYIT-TCQPSWSVYKNIERFRKTYQELFFPHEDNEKDLPISKLIEEDAQLTQMEDV 194
Db 121 MKFYARVCRSGSLGVGHQLEEFLLNQSSPTYFINGDR-----IDSLLENDRQOHTVLDV 174
QY 195 FSQLTVDVNSLFRNSFNVFRMQOQEDQTFQSHFTSDTLTPEYFPFAPSKPEMTKADL- 253
Db 175 -----MQDSFDRASSI-----MDLFPQDRFT-REPQDTYISYSPFS-SPHRRSLL 218
QY 254 -----ECWDIP-----NFFQLECNFSVSIYSESVSETITKMLKAIEDLPKQD--KAP 298
Db 219 FNPKSRFARNIMHFPYRHLFNDMPQPFDMHQAAQAMNLHLRLPDQLPMTFEFSGD 278
QY 299 DHGGLISKMLPGDQGLGELQNLRSRCFEKHEKCKQCOAHLSEDC-----PDVPAHTEL 354
Db 279 NH-----DRTVCKEIRHNSTGCLMKKDKQCEKCEILLVDCSTNNPSQMLQREL 327
QY 355 DEAIRLVNWSNOQYGOIQMLTKRKHLEDTAYLVKRGVQFGWVSELANQAPETEIFNSIQ 414
Db 328 NNSLQLAEEKFTKLYDELLQSYQEKMLNTSSLKQLNEQFSWVSQLANLTQGEDQYYIQLVT 387
QY 415 VVPIRIEHNISKODETMMTDLSLTSSNFTLKPILEESAESSNFTGYGVAKALQHFKE 472
Db 388 TVSS-HNSDSEVPVGUTRVVVVLLFQSYPTVTVPEVSRNNPKFMETVAERALQEYRQ 444

```

QY 432 MTDLSILPSSNFTLKIPLEESAESSNFGYVYVAKALQHFKH 473  
DB 404 VT-VQFDFSEPLSLVPGDISWDDPRFWEIVAEQALQHYKON 444

## RESULT 6

Q9JK98 PRELIMINARY; PRT; 372 AA.  
AC Q9JK98;  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE CLUSTERIN ISOFORM 2.  
OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.  
OX NCBI\_TaxID=10090;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BALB/C; TISSUE=UTERUS;  
RA You K.H., Jeon J.H.;  
RT "Identification of Truncated SGP-2 Lacking a Signal Peptide for  
RT Secretion.";  
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH (BY SIMILARITY).  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
DR EMBL; AF248058; AAF67185.1; -.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Glycoprotein.  
SQ SEQUENCE 372 AA; 43109 MW; A1047A7E48520FEE CRC64;

Query Match 12.7%; Score 321; DB 11; Length 372;  
Best Local Similarity 23.6%; Pred. No. 2.2e-14;  
Matches 93; Conservative 83; Mismatches 170; Indels 48; Gaps 12;

QY 95 KQALKLLNEVQBHEERLCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVYKNI 153  
DB 4 REDALETRDSEMKLAFPEVCNETHMALWEECKPCLKHTCKMFKYARVCRSGSLVGQQL 63  
QY 154 ERFERKIYQFLFPFHEDNEKDLPISKELTEDDAQLTQ----MEDVFSQLTVDVNSLFRNS 209  
DB 64 EEFNQSSPPYFWMGDR-----IDSLLESDRQSQVLDAQMSDFRSGIIDLTFQDR 117  
QY 210 FNVFRMQQEFDTFQSHFISDLDLTPYFFP-----AFSKEPMTKADLEC-NDIPNF 262  
DB 118 F--FAR-----ELHDPHYFS-----PIGFPKRPHEFLYPKSLRSLMSFSHYGPPSF 163  
QY 263 FQLCFNFSYIYSVSETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ-DRGLCGELDQ 321  
DB 164 HNMFPFFEMTHQAQAMDVQLHSPAFQPPD-----FLREGEDDRTVCKEIRR 213  
QY 322 NLSRCFKFHEKCKQKQAHLSDC-----PDVPALHTELDEAIRLVNVSNOQYGOILQWTRK 377  
DB 214 NSTGCLMKQCKEKOBIILSDVDCSTNPAQANLRQELNDSLQVAERLTQYKELLOSFOFS 273  
QY 378 HLEDATYLVKMRGQFQWSELANQAPETIIFNSIQVVPRIHEGNISKQDEMTMDLSI 437  
DB 274 KMLNTSLLLEQLNDQFNWSQLANLTQGEDKYLYRSTV-TTHSSDSEVPSTVEVVVKL 332  
QY 438 LPSNFTLKIPLEESAESSNFGYVYVAKALQHFK 471  
DB 333 FDSDPITVVLPEEYVKDNPKFMDTVAEKALQEYR 366

## RESULT 7

Q35510 PRELIMINARY; PRT; 295 AA.  
AC Q35510;  
DT 01-JAN-1998 (Tremblrel. 05, Created)  
DT 01-JAN-1998 (Tremblrel. 05, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE CLUSTERIN.  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=WISTAR; TISSUE=SEMINAL VESICLE;  
RX MEDLINE=98163873; PubMed=9503143;  
RA Izawa M.;  
RT "Identification of a transcript predicting an alternative form of  
RT sulfated glycoprotein-2 (clusterin) in rat tissues.";  
RL Biochem. Mol. Biol. Int. 44:9-19(1998).  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH (BY SIMILARITY).  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY  
CC SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
DR EMBL; D11339; BAA21775.1; -.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Glycoprotein.  
SQ SEQUENCE 295 AA; 34115 MW; C1E07FA745DD5470 CRC64;

Query Match 10.5%; Score 266.5; DB 11; Length 295;  
Best Local Similarity 25.2%; Pred. No. 9.2e-11;  
Matches 80; Conservative 59; Mismatches 131; Indels 47; Gaps 11;

QY 115 LCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVYKNIERFKIYQFLFPFHEDNEK 173  
DB 9 VCNETHMALWEECKPCLKHTCKMFKYARVCRSGSLVGRLQLEEFNQSSPFFFWMGDR-- 66  
QY 174 DLPISKELTEDDAQLTQ----MEDVFSQLTVDVNSLFRNSFNFRMQQEFDTFQSHFI 229  
DB 67 ----IDSLLESDRQSQVLDAQMSDFRSGIIDLTFQDRF--FTHEPQDI-----HHF- 114  
QY 230 SDTDLTEPYFFP-----AFSKEPMTKADLECQNDIP-NFQQLCFNFSYIYSVSETIT 282  
DB 115 -----SPMGFPKRPHEFLYPKSLRSLRSLMPLSHYGLPSLFHNMFPQPFEDMTHQQAAMDV 168  
QY 283 KMLKAIEDLPKQKAPDHGGLISKMLPGQ-DRGLCGELDQNLRCFKFHEKCKQKQAHLS 341  
DB 169 QJHSPALQFPDQVDFLKE-----GEDDPTVCKEIRHNSTGCLMKMGQCKQCEILS 218  
QY 342 EDC-----PDVPALHTELDEAIRLVNVSNOQYGOILQWTRKHLDTAYLVKMRGQFGWVS 397  
DB 219 VDCSTNPAQANLRQELNDSLQVAERLTQYVNEILLSQSKMLNTSLLEQLNDQFSWVS 278  
QY 398 ELANQAPETIIFNSIQ 414  
DB 279 QLANTLQGGDDQYFGSPQ 295

## RESULT 8

Q9ERD1 PRELIMINARY; PRT; 218 AA.  
ID Q9ERD1  
AC Q9ERD1;

01-MAR-2001 (Tremblrel. 16, Created)  
01-MAR-2001 (Tremblrel. 16, Last sequence update)  
01-DEC-2001 (Tremblrel. 19, Last annotation update)  
CLUSTERIN (FRAGMENT)  
Rattus norvegicus (Rat)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
NCBI\_TaxID=10116;  
[1]  
SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE DAWLEY; TISSUE=THYMUS;  
RA Park J.H., Park J.S., Ju S.K., Na S.Y., You K.H.:  
RT "Determination of clusterin mRNA expression of apoptosis induced rat  
thymocytes in vivo and in vitro."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH (BY SIMILARITY).  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY  
SIMILARITY).  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
DR EMBL; AF314657; AAG31162.1; -  
DR InterPro: IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 2.  
DR SMART; SM00035; Clu; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Glycoprotein.  
FT NON\_TER  
SQ SEQUENCE 218 AA; 24931 MW; ED2FC4425A510589 CRC64;

Query Match 8.1%; Score 204; DB 11; Length 218;  
Best Local Similarity 28.3%; Pred. No. 1.2e-06;  
Matches 47; Conservative 36; Mismatches 73; Indels 10; Gaps 3;

QY 312 DRGLGELDONLSRCPKFEKQKQAHLSQEDC-----PDVPALHTLDEAIRLVNVSNOQ 367  
DB 51 DRTVCKEIRNHTGCLMKRGQCEKQELLSDVCSNTPAQANLRQLNDSLQVAERLTQQ 110  
QY 368 YGQILQWTRKHLEDYALVEKMRGQGVWSELAN--QAPETELIFNSIQVVPRIHEGNIS 425  
DB 111 YNELLHSLGSKMLNTSLLEQLNDQFSWVSQLANLTQGGDDQYLRVSTVTT-----HSSDSE 166  
QY 426 KQDETMTDLISLPSSNFTIKPLEESAESSNFIGYVVAKALQHFK 471  
DB 167 VPSRVTEVVVKLFDSPITVWLPPEVSKDNPKFMDTVAEKALQEYR 212

RESULT 9  
O76329 ID O76329 PRELIMINARY; PRT; 1738 AA.  
AC O76329;  
DT 01-NOV-1998 (Tremblrel. 08, Created)  
DT 01-NOV-1998 (Tremblrel. 08, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE INTERAPTIN.  
GN ABPD.  
OS Dictyostelium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyostelida; Dictyostelium.  
OX NCBI\_TaxID=44689;  
[1]  
SEQUENCE FROM N.A.  
RX MEDLINE=98365468; PubMed=9700162;  
RA Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;  
RT "Interaptin, an actin-binding protein of the alpha-actinin superfamily  
in dictyostelium discoideum, is developmentally and cAMP-regulated and  
RT associates with intracellular membrane compartments.";  
RL J. Cell Biol. 142:735-750(1998).  
DR EMBL; AF057019; AAC34582.1; -  
DR HSSP; 001082; 1BKR.  
DR InterPro: IPR001589; Actinin\_act\_bind.  
DR InterPro: IPR001715; Calponin\_hom.

DR InterPro: IPR001990; Granin.  
DR InterPro: IPR001451; Hexapep\_transf.  
DR InterPro: IPR002017; Spectrin.  
DR Pfam; PF00307; CH; 2.  
DR SMART; SM00033; CH; 2.  
DR PROSITE; PS00019; ACTININ\_1; UNKNOWN\_1.  
DR PROSITE; PS00020; ACTININ\_2; 1.  
DR PROSITE; PS00021; CH; 2.  
DR PROSITE; PS00422; GRANINS\_1; UNKNOWN\_1.  
DR PROSITE; PS00101; HEXAPEP\_TRANSFERASES; UNKNOWN\_1.  
SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Query Match 6.3%; Score 160; DB 5; Length 1738;  
Best Local Similarity 20.9%; Pred. No. 0.015;  
Matches 102; Conservative 88; Mismatches 171; Indels 128; Gaps 21;

QY 36 KDKTATSENL-----KSFSEVGEDADEEVKALTGKQKIMMER----- 76  
DB 667 KDNQTINEQNLKOLSEKDEKLEKLSNQEQQDEKINNLLLEIKEDCLTERINQOLLEN 726  
QY 77 -----KEKEHTNLMSTLKKCREKQKALKLLNEVQEHLEEEER 114  
DB 727 IDLSKYQQLLEFEFNKLSKE-----NOLNELQSKQDERFNOLND--EKLEKEKQ 779  
QY 115 LCRESLADSWGECRSCLENNCMRIYTTTCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKD 174  
DB 780 L--QSIEDEFNQY-----QQQLSSNSN-IDQQLQSTIIELSELKEQKELN 822  
QY 175 LPISEKLIBEDAQLTOMEDVFSOLT-----VDVNSLFNRSFNVRQMQQEFDTQFSH 227  
DB 823 ---DSKLIBEKEKLOOQOEFDQLEKNEKNKHODQLELLEKQ---LKQLQOEYDQLNETN 876  
QY 228 FISTDTLTPEYFPFAPSKPEMTKADLE-----QCHWDIPNFFQFCNFSVSIYESV 277  
DB 877 QSTENOLNQNLT---INKENLNEKEQELLKLQNLNQKIEKIQFDQOEFSKQNSINIELV 933  
QY 278 SETITKMLKAIED---LPKQDKA---PDHGLISKMLPGQDRGLCGELDQNLNLSRCFKFHE 331  
DB 934 NEKNEKLIQLOQDYDOLKQONRSNDEKENDLLEK--ENOLKSIQNELNOLIEKNSDVK 991  
QY 332 KQKQCAHLSQEDCPD---VPALHTLDE----AIRLVNVSNOQYQOIL-----Q 373  
DB 992 EQOLKQOQSIENDLIEKENQIQLOQSLNEQROQOQSNQSEKDOQLNOLIEKNQDOKEQ 1051  
QY 374 MTEKHLEDYATY----LVEKMRGQFG-WVSELANQAPETELIFNSIQVVPRIHEGNISKQD 428  
DB 1052 LKQOQSIENDLFEKENQIQLOQSLNEQROQOQSNQSEKDOQLNOLIEKNSDVK 1105  
QY 429 ETMTDLISI 437  
DB 1106 EQOLKQOQSI 1114

RESULT 10  
Q9W6V0 ID Q9W6V0 PRELIMINARY; PRT; 3616 AA.  
AC Q9W6V0;  
DT 01-NOV-1999 (Tremblrel. 12, Created)  
DT 01-NOV-1999 (Tremblrel. 12, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE HYPERION PROTEIN, 419 KDA ISOFORM.  
GN HYPERION.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
[1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=RETINA;  
RA Kemmer W.A., Schwarz U.;  
RT "Characterization of Hyperion, a gene coding for an abundance of gene





```
DR InterPro: IPR001849; PH.
DR InterPro: IPR000961; PKinase_C.
DR InterPro: IPR000861; REM_repeat.
DR InterPro: IPR002290; Ser_thr_pkinase.
DR Pfam: PF00130; DAG_PE_bind; 1.
DR Pfam: PF02185; HR1; 1.
DR Pfam: PF00169; PH; 1.
DR Pfam: PF00069; PKinase; 1.
DR SMART: SM00109; C1; 1.
DR SMART: SM00074; HR1; 1.
DR SMART: SM00233; PH; 1.
DR SMART: SM00220; S_TKC; 1.
DR SMART: SM00133; S_TKC_X; 1.
DR PROSITE: PS50081; DAG_PE_BIND_DOM_2; 1.
DR PROSITE: PS50003; PH_DOMAIN; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; 1.
DR PROSITE: PS50011; PROTEIN_KINASE_DOM; 1.
DR PROSITE: PS00108; PROTEIN_KINASE_ST; 1.
DR KW ATE-binding; Kinase; Serine/threonine-protein kinase; Transferase.
SQ SEQUENCE 1388 AA; 160799 MW; CAB3CE7D3860465D CRC64;

Query Match 5.7%; Score 145; DB 6; Length 1388;
Best Local Similarity 16.5%; Pred. No. 0.12; 166;
Matches 94; Conservative 97; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISNLSKFSVEGIDEAEVVKALTGKIKOMKI-----MMERKEKEHTNLM 86
DB 780 KOKVLNEDVRNLT-----LKIEQETQRCLTQNDLKMOTQOVNTLKMSEKQLKQENHLL 835
QY 87 TLK-----KCKEKEQALKLNEVQEHLEEE-----RCLR 117
DB 836 EKMSLEKQNALRKERDAGQCKELQDQLEAQYFSTLYKTVRELKECEKTKLCK 895
QY 118 E-----SLADSWGECRSCLNENCMRIYTTQCPSSWSSVKNKIERFFRKIYFLFPFHEDN 171
DB 896 ELQKKQELQDE-----RDSLAQLETLTKADSEQLARSIAEQSDLEKE 942
QY 172 E--KDLPISE-----KLIEDAQLTQMEDVFSQLTVDVNSLNFNSFNFRQMQQBFQD 222
DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLSDVANLANEKEELNKLKEAQEQ 1002
QY 223 TFQSHFISDTLTPEYFPFAPFSKPMTKADLEQCWDIPNFFQFCNFSVSIYESVSETIT 282
DB 1003 LSR---LKDEISAAAKAQEKOLLERTLK-----TQAVN 1036
QY 283 KMLKAIE-----DLPKQKAPDHGGLISKMLPGDGRGLGEL-----DONLSRCF 327
DB 1037 KLABIMNRKEPVKRGNDTDVRRKEK-----ENRKLHMLKSEREKLITQOMI 1082
QY 328 KFHEKCKCOAHLSEDCPDVPALHTLDEAIRLVNVSNOQYQGLTQWKHLE-----380
DB 1083 KYQKELNEMQAIQAE-----SQRIEQLMTLDSKSDIEQLRSQQLALHGLDSSSIGSG 1138
QY 381 -----DTAYLVEKMRG-----QFGWVSEL-----399
DB 1139 PGDTEADGGPSPSEGLWLSLPVNRNNTKFGWKYIVVSSKTLFLFYDSQDQKEQSNPYM 1198
QY 400 -----ANQAPETEIIFNSIQVVPRI-----HEGNTSKQDETMTDLTSLPSSNFT 444
DB 1199 VLDIDKLFHVRVPVTDVYRADAKEIRIFQILYANEGSKKEQ-----1243
QY 445 LKIPLEESAESSNFI---GYVAKALQHF 470
DB 1244 --FPVEPVGEKSNYCHKGHEFTPLHYF 1270

RESULT 15
Q9FJ35
ID Q9FJ35 PRELIMINARY; PRT; 1305 AA.
AC Q9FJ35;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
```

```
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE MYOSIN HEAVY CHAIN-LIKE PROTEIN.
OS Arabidopsis thaliana (Mouse-ear cress).
OC Eukaryota; Viridiplantae; Streptophyta; Tracheophyta;
OC Spermatophyta; Magnoliophyta; eudicotyledons; Core eudicots; Rosidae;
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=COLUMBIA;
RX MEDLINE=99087489; PubMed=9872454;
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,
RA Tabata S.;
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.
RT Sequence features of the regions of 1,013,767 bp covered by sixteen
RT physically assigned P1 and TAC clones.";
RL DNA Res. 5:297-308(1998).
DR EMBL; AB016871; BAB10654.1;
DR ENSEMBL;
SQ SEQUENCE 1305 AA; 149961 MW; F37284B8E7BBF217 CRC64;
```

Query Match 5.7%; Score 144; DB 10; Length 1305;  
Best Local Similarity 18.8%; Pred. No. 0.14;  
Matches 94; Conservative 82; Mismatches 176; Indels 148; Gaps 18;

```
QY 26 LKDSHCAPTWKDKTAISENLKFSFSEGEIDAEVVKALTGKIKOMKIMMERKEKEHTNLM 85
DB 127 LKESHV---KERLFS---LRDIHEIHQRDSS-----TRASELAQLESSKQOVSDLS 174
QY 86 STLKKKREEKQEAULKNEVQEHLEEBERLCRESLA-----DSWGECRSCLNENCMRIY 139
DB 175 ASLKAABEENKAISKNVETMKNLEQTQNTIQELMAELGKLKDSHREKESLS-----228
QY 140 TTCOPSSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIEED-----AQLTQM 191
DB 229 -----LVEVHETHQRDSSIHVKELEEEQVESKKLVAELNQT 264
QY 192 -----EDVFSQLTVDVNSLNFNSFNFRQMQQFQDTFQSHFISDTLTPEYFPFAPS 244
DB 265 LNNAEEKKVLKSQIAELSNELKEAONTIQELYSESQKLESVSKDRDL-----FS 316
QY 245 -----KEPMTK-ADLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKQ 294
DB 317 LRDIETHQRSSTRVSELEA-----QL-----ESSQRISDLTVDLKDAEEE 359
QY 295 DKAPDHGGLISKMLPGDGRGLGELDONLSRCFKFHEKCKCOAHLSEDCPDVPALHTEL 354
DB 360 NKA-----ISSKNLEIMDK-----LEQAONTIKELMDELGELKDRHKEK 398
QY 355 D-BAIRLVNVSNOQYQGLTQWKHLEDYALVVEKMRGQFGWVSELANAQAPET-----406
DB 399 ESELSSLVKSADQOVADMKQSL-----DNAEEKKMLSQ--RILDISNEIQEAQKTIQEH 451
QY 407 -----EIIFNSIQVVPRIHEGNISKQDETMTDLTSLPSSNFTKIPLEES 452
DB 452 MSSEQLKESHGKVERELTGLRDIETHQRSSTRVSELETLQKLEQRVVDLSASLNAA 511
QY 453 AESSNFIQYVVAKALQHFKE 472
DB 512 EEKKSILSSMILEITDELQK 531
```

Search completed: July 2, 2002, 11:49:57  
Job time: 391 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:51:19 ; Search time 56.51 Seconds  
(without alignments)  
937.572 Million cell updates/sec

Title: US-09-722-544A-4MOD  
Perfect score: 2529  
Sequence: 1 MRTWYSNGNKKPLLVFI.....FIGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2485	98.3	477	AA130786	Protein encoded by
2	2460	97.3	466	AA130793	Immature human HKN
3	2437	96.4	495	AA130785	Protein encoded by
4	2346	92.8	446	AA130792	Mature secreted hum
5	1753.5	69.3	465	AA130791	Bovine HKNG1 ortho
6	1635	64.7	466	AA130787	Protein encoded by
7	1554	61.4	450	AA130788	Guinea pig HKNG1 o
8	1277	50.5	374	AA130789	Guinea pig HKNG1 o
9	1272.5	50.3	373	AA130790	Guinea pig HKNG1 o
10	780.5	30.9	521	AA130794	Amino acid sequenc
11	434	17.2	449	AA128048	Novel human secret

12	432	17.1	448	12	AA111704	Cytolysis Inhibito
13	405.5	16.0	416	22	ABB50285	Apolipoprotein J o
14	400.5	15.8	446	21	AA130341	Porcine clusterin
15	320	12.7	363	22	AA130364	Human gene 1 encod
16	173.5	6.9	148	22	AA130367	Human polypeptide
17	160.5	6.3	116	21	AA130375	Human secreted pro
18	159.5	6.3	139	22	AA130383	Human gene 1 encod
19	159.5	6.3	139	22	AA130383	Human gene 1 encod
20	152.5	6.0	247	22	AA130387	Human gene 4 encod
21	149	5.9	1372	19	AA156473	Human gene 1 encod
22	145	5.7	944	21	AA167600	Protein with Rho p
23	145	5.7	1388	19	AA156475	Protein with Rho p
24	143	5.7	1931	22	AA161012	Drosophila melanog
25	141.5	5.6	3433	18	AA122017	Utrrophin. Homo sa
26	140	5.5	934	22	AA101768	Human secreted pro
27	140	5.5	2517	21	AA171159	Human phosphodiester
28	139.5	5.5	2688	22	AA140883	Human polypeptide
29	139	5.5	2633	22	AA106505	Novel human diapo
30	138.5	5.5	2663	22	AA139097	Human polypeptide
31	138	5.5	5373	22	AA146603	Novel bone marrow
32	138	5.5	5447	22	AA146697	Novel bone marrow
33	135.5	5.4	1427	12	AA110534	Human 160kD mediat
34	135	5.3	1374	22	AA169070	Human male enhance
35	133	5.3	990	22	AA178520	Human protein SEQ
36	131.5	5.2	1411	17	AA102258	Nucleolar/endosoma
37	130.5	5.2	2482	16	AA172826	Human mitosis. Ho
38	130.5	5.2	2482	19	AA123996	Human mitosis amin
39	130	5.1	1788	22	AA140467	Human polypeptide
40	130	5.1	3248	17	AA199795	Kinetochore protei
41	129	5.1	931	22	AA179504	Human protein SEQ
42	128.5	5.1	1392	20	AA106999	Restin protein seq
43	127.5	5.0	721	21	AA121227	Protein encoded by
44	127.5	5.0	2053	22	AA103501	Human protein kina
45	127	5.0	1047	18	AA101535	Cellular homologue

ALIGNMENTS

RESULT 1  
ID AAY30786 standard; Protein; 477 AA.  
AC AAY30786;  
DT 23-NOV-1999 (first entry)  
XX Protein encoded by a human HKNG1 splice variant HKNG1-V1.  
DE HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX Homo sapiens.  
XX WO9947535-A1.  
PN 23-SEP-1999.  
XX 16-MAR-1999; 99WO-US05606.  
XX 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
XX 22-JAN-1999; 99US-0236134.  
PA (MILL-) MILLENNIUM PHARM INC.  
PA (REGC) UNIV CALIFORNIA.  
XX Chen H, Freimer NB;  
XX WPI; 1999-562047/47.  
DR N-PSDB; AA210751.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
XX Claim 1; Fig 2A-B; 205pp; English.  
XX  
XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)  
XX splice variant. HKNG1 is a gene associated with bipolar affective  
XX disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
XX modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
XX activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
XX activity in individuals can then be administered therapeutically to  
XX treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
XX e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
XX early-onset autosomal dominant myopia. The polynucleotides can be used  
XX in gene therapy techniques to treat such disorders. They are also useful  
XX in diagnosis to identify individuals having, or at risk of developing,  
XX HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
XX mutations especially result in the production of a protein with a  
XX different sequence to the human full-length HKNG1 polypeptide or  
XX splice variant sequences, especially the substitution of a lysine for  
XX a glutamic acid at residue 202 or 184. The polynucleotides are also  
XX useful in gene mapping, to produce probes or primers to identify  
XX similar sequences (e.g. mutants or sequences from different species)  
XX and to produce transgenic animals.  
XX  
SQ Sequence 477 AA;

Query Match 98.3%; Score 2485; DB 20; Length 477;  
Best Local Similarity 99.2%; Pred. No. 4.8e-201;  
Matches 472; Conservative 1; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRTDWSNGNKPPLLYFIVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEV 60  
Db 1 mrtdwysngmkkppllyfvicllwkshcapwkdktaisenlksfsevgeldadeev 60  
Qy 61 KKAALGFIKQMTIMRKEKEHNTLMSTLKKCREEKOALKLNEVOEHLEEBEERLCRESL 120  
Db 61 kkalgtikqmkimrkekehtnlmstlkkcreekqealkllnevqehleeeerlcsesl 120  
Qy 121 ADSWGECSLENNCMRYTTTCOPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEK 180  
Db 121 adswgecsleenncmrlytttcopsswssvknkierfrkiyqflfpfhednekdipisek 180  
Qy 181 LIEKDAQLTQMEDVFSQLTVDVNSLFNSFNFRMQQFEEDQTFQSHFISDLDLPEYFF 240  
Db 181 lieedaqltqmedvfsqldtdvnsifnfrmqgqefdtqskfsiddltpyff 240  
Qy 241 PAFSKEPMTKADLEQWDIPNFQFCNFSVSIYSEVSETITKMLKATEDLPKQKAPDH 300  
Db 241 pafskepmtkadleqwdipwffqfcnfsvsiyesevsetitkmlkatedlpkqkadh 300  
Qy 301 GGLISKMLPGDQGLCGELDONLSRCFHFKEKQCOAHLSDCQDPVPALHTLDEAIRL 360  
Db 301 ggllskmlpgdqrqglcgeidnlsrccfkmekccqahlsdcqdpvpalhteldeairl 360  
Qy 361 VNVSNQYQGIQLQMTKRKLEDAYLVEKMRGQFGWSELANOAPETETIFNSIQVVRH 420  
Db 361 vvnvsnqyqgiqlqmrkrkledtaylveknrgqfgwselanqapetetiifnsiqvvrh 420  
Qy 421 EGNISKQDTMTDLSILPSSNFTLIKIPLEESAESSNFYGVVAKALQHFKEHFT 476  
Db 421 egniskqdetmttdlsilpssnftlikipleesaessnfgyvvakalqhfkehft 476

RESULT 2  
AAY30793  
ID AAY30793 standard; Protein; 466 AA.  
XX  
AC AAY30793;  
XX

DT 23-NOV-1999 (first entry)  
XX  
DE Immature human HKNG1 protein form 2.  
XX  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
XX 16-MAR-1999; 99WO-US05606.  
XX  
XX 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Chen H, Freimer NB;  
XX  
DR WPI; 1999-562047/47.  
XX  
XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
PS Claim 22; Fig 17; 205pp; English.  
XX  
XX The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1  
CC is a gene associated with bipolar affective disorder (BAD). HKNG1  
CC polynucleotides are useful to identify compounds modulating HKNG1 gene  
CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting  
CC or enhancing HKNG1 gene expression or activity in individuals can then  
CC be administered therapeutically to treat HKNG1-mediated disorders,  
CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or  
CC HKNG1-mediated myopia disorders, such as early-onset autosomal  
CC dominant myopia. The polynucleotides can be used in gene therapy  
CC techniques to treat such disorders. They are also useful in diagnosis  
CC to identify individuals having, or at risk of developing, HKNG1-mediated  
CC disorders due to mutations in the HKNG1 gene. Such mutations especially  
CC result in the production of a protein with a different sequence to  
CC the human full-length HKNG1 polypeptide or splice variant sequences,  
CC especially the substitution of a lysine for a glutamic acid at residue  
CC 202 or 184. The polynucleotides are also useful in gene mapping, to  
CC produce probes or primers to identify similar sequences (e.g. mutants  
CC or sequences from different species) and to produce transgenic  
CC animals.  
XX  
SQ Sequence 466 AA;

Query Match 97.3%; Score 2460; DB 20; Length 466;  
Best Local Similarity 99.8%; Pred. No. 5.9e-199;  
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 12 MKPPLLVFTVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKALTGIKQMK 71  
Db 1 mkppllvfvicllwkshcapwkdktaisenlksfsevgeldadeevkaltgikqmk 60  
Qy 72 IMMERKEHNTLMSTLKKCREEKOALKLNEVOEHLEEBEERLCRESLADSWGCRSL 131  
Db 61 immerkehtnlmstlkkcreekqealkllnevqehleeeerlcsresladswgcrsl 120  
Qy 132 ENNCMRIYTTTCOPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKDLQLTQM 191  
Db 121 enncmriytttcopsswssvknkierfrkiyqflfpfhednekdplisekdlqldtqm 180



QY 192 EDVFSQTLVDVNSLFRNSFNVRQMOQEFDTQFSHFISDTDLTEPYFFPAFSKEPMTKA 251  
 Db 181 edvfgltvdvnslnfrsnfvrmqgefqtfgshfisdtdltepyffpafskemtk 240  
 QY 252 DLEQCDWIPNFFQFCNFSVSYESVETITKMLKAIEDLPKODKAPDHGGLISKMLPGQ 311  
 Db 241 dleqcdwipnffqfcnfsvsyessvettitkmlkaiedlpkdkapdhggliskmlpgq 300  
 QY 312 DRGLGELQDLNLSRCFKFHEKCKOAHLSDECDPVPALHTELDLRLVNVSNQYGOI 371  
 Db 301 drglcgelqdlnlsrcfkfhekkckqahlsedcdpvpalhlteldlrlvvnsgqygi 360  
 QY 372 LQMKRKHLEDAYLVKMRGQFGWSELANQAPETETIIFNSIQVVPRIHEGNISKQDET 431  
 Db 361 lqmktrkhledtaylvkrmrgqfgwswselanqapetetiifnsiqvvprihegniskqdet 420  
 QY 432 MTDLSLPSNNTLTKTLPLESAESSNFYGVVAKALQHFKEHFTW 477  
 Db 421 mtdslpsnntltktlplesaesnsnfygyvvakalqhfkehftw 466

RESULT 3  
 AAY30785  
 ID AAY30785 standard; Protein: 495 AA.  
 XX  
 AC AAY30785;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Protein encoded by human HKNG1 cDNA.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 OS Homo sapiens.  
 XX  
 XX WO9947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-US05606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 PR 05-JUN-1998; 98US-0088312.  
 PR 28-OCT-1998; 98US-0106056.  
 PR 22-JAN-1999; 99US-0236134.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PA (REGC) UNIV CALIFORNIA.  
 XX  
 PI Chen H, Freimer NB;  
 XX  
 DR WPI; 1999-562047/47.  
 DR N-PSDB; AAZ10750.  
 XX  
 PT New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia  
 XX  
 PS Claim 1; Fig 1A-B; 205pp; English.  
 XX  
 CC The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1  
 CC is a gene associated with bipolar affective disorder (BAD). HKNG1  
 CC polynucleotides are useful to identify compounds modulating HKNG1 gene  
 CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting  
 CC or enhancing HKNG1 gene expression or activity in individuals can then  
 CC be administered therapeutically to treat HKNG1-mediated disorders,  
 CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or  
 CC HKNG1-mediated myopia disorders, such as early-onset autosomal  
 CC dominant myopia. The polynucleotides can be used in gene therapy  
 CC techniques to treat such disorders. They are also useful in diagnosis  
 CC to identify individuals having, or at risk of developing, HKNG1-mediated

disorders due to mutations in the HKNG1 gene. Such mutations especially  
 result in the production of a protein with a different sequence to  
 the human full-length HKNG1 polypeptide or splice variant sequences,  
 especially the substitution of a lysine for a glutamic acid at residue  
 202 or 184. The polynucleotides are also useful in gene mapping, to  
 produce probes or primers to identify similar sequences (e.g. mutants  
 or sequences from different species) and to produce transgenic  
 animals.

XX  
 SQ Sequence 495 AA;  
 Query Match 96.4%; Score 2437; DB 20; Length 495;  
 Best Local Similarity 98.7%; Pred. No. 5.6e-197;  
 Matches 464; Conservative 2; Mismatches 4; Indels 0; Gaps 0;

QY 7 SNGSNKPPPLVFIYCLLWLDKSHCAPTWKDKTATSENLSKSFSEYGEIDADEVKKALTG 66  
 Db 25 nsgsnkppplvfiycllwlldkshcawtwdkdtatse nlsksfsevegeidadevkkaltg 84  
 QY 67 IKQKIMMERKEKHTNLMSTLKKCREKQKALLLNEVQHELEEEERLCRESLADSGE 126  
 Db 85 ikqkimmerkekehtnlmstllkkcreekqalkllneveqheleeeerlcresladshg 144  
 QY 127 CRSCLENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLFPHPHEDNEKDLPISEKLEKDA 186  
 Db 145 crsclenncmriyttccpswssvknkierfrkiyqflfphephednekdlpiseklekda 204  
 QY 187 QLTQMEDVFSQTLVDVNSLFRNSFNVRQMOQEFDTQFSHFISDTDLTEPYFFPAFSKE 246  
 Db 205 qltqmedvfgltvdvnslnfrsnfvrmqgefqtfgshfisdtdltepyffpafsk 264  
 QY 247 PMTKADLQCDWIPNFFQFCNFSVSYESVETITKMLKAIEDLPKODKAPDHGGLISK 306  
 Db 265 pmtkadlegcdwipnffqfcnfsvsyessvettitkmlkaiedlpkdkapdhgglisk 324  
 QY 307 MLPQDGRGLCGELQDLNLSRCFKFHEKCKOAHLSDECDPVPALHTELDLRLVNVSNQ 366  
 Db 325 mlpgdgrlgcelqdlnlsrcfkfhekkckqahlsedcdpvpalhlteldlrlvvnsg 384  
 QY 367 YGQITLQMKRKHLEDAYLVKMRGQFGWSELANQAPETETIIFNSIQVVPRIHEGNISK 426  
 Db 385 ygqitlqmktrkhledtaylvkrmrgqfgwswselanqapetetiifnsiqvvprihegnisk 444  
 QY 427 QDETMMKTDLSLTPSSNFTLKIPLESAESSNFYGVVAKALQHFKEHFTW 476  
 Db 445 qdetmmtldslpsnntltktlplesaesnsnfygyvvakalqhfkehft 494

RESULT 4  
 AAY30792  
 ID AAY30792 standard; Protein: 446 AA.  
 XX  
 AC AAY30792;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Mature secreted human HKNG1 protein sequence.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 OS Homo sapiens.  
 XX  
 XX WO9947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-US05606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 PR 05-JUN-1998; 98US-0088312.



Db 1 mkppllvfiyqlldrcqcapgtgkrtisredpqgfskageidvdeevkkaaligmkqm 60  
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOEHLEERLCRESLADSGWGERSC 131  
 Db 61 ilmerreeehsklmrtllkckreekqalklmnevqehleeeerlcqvsimgvdeckscl 120  
 QY 132 ENNCRIYTTTCOPSSWSVKNKIERFRKIKYQFLFPFHEONEKDLPISEKLEKDAOLTM 191  
 Db 121 esdcmrfyttccqsswsmkstiervrkikyqflfpfheddekelvgekteedvqlmqi 180  
 QY 192 EDVFSQTLVDVNSLFRNSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251  
 Db 181 envfsqtlvdvgylnmsfsvfkmgqgdflafqsfmsdtdsmeyffpafskapakka 240  
 QY 252 DLQOCWDIPNFFOLFNCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311  
 Db 241 hmqswdipffqlfncfslsvygsatvtemikaedlskqdsahggpsstcpvr 300  
 QY 312 DRGLCGELDONLSRCFKHEKQKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYGOI 371  
 Db 301 grglcgepgqnsseciqfharccqkcqylwadcpavpelytkadealelvnlsnqyqav 360  
 QY 372 LQMTREKHELTAYLVEKMRGQFCGWSELANOAPETETIIFNSIQVVPRIHEGNISKODETM 431  
 Db 361 lqmtqhlhdedtlylmekmregfgwvteiasqtpgseniifsklvvpgvhegnfskqdkm 420  
 QY 432 MTDLSILPSSNFTLKIPLESSESSNFYGVYVAKALQHEKHEFTW 477  
 Db 421 i-disilpsnftltipleesessdfisymiakavqhkhfkfw 465

RESULT 6  
 AAY30787  
 ID AAY30787 standard; Protein; 466 AA.  
 XX  
 AC AAY30787;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Protein encoded by the guinea pig HKNG1 ortholog gphkng1815.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.

OS Cavia cobaya.

PN WO947535-A1.

PD 23-SEP-1999.

PF 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

PA (MILL-) MILLENNium PHARM INC.

PA (REGC) UNIV CALIFORNIA.

PI Chen H, Freimer NB;

DR WPI; 1999-562047/47.

DR N-PSDB; AAZ10755.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of

PT neuropsychiatric disorders, e.g. bipolar affective disorders and

PT schizophrenia

PS Claim 1; Fig 7A-B; 205pp; English.

XX

CC The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.  
 CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective  
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
 CC activity in individuals can then be administered therapeutically to  
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 466 AA;

Query Match 64.7%; Score 1635; DB 20; Length 466;  
 Best Local Similarity 66.5%; Pred. No. 2,3e-129;  
 Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY 12 MKPPLLVFIYQLLDRCQCAPGTGKRTISREDPQGFSGKAGEIDVDEEVKKAALIGMKQM 71  
 Db 1 mkppllmfpvcllwldkchcpcaptwkdtaiensanufseageidvdeevkkaaligkmk 60  
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOEHLEERLCRESLADSGWGERSC 131  
 Db 61 ilmerreeehsklmrtllkckreekqalklmnevqehleeeerlcqvsimgvdeckscl 120  
 QY 132 ENNCRIYTTTCOPSSWSVKNKIERFRKIKYQFLFPFHEONEKDLPISEKLEKDAOLTM 191  
 Db 121 esdcmrfyttccqsswsmkstiervrkikyqflfpfheddekelvgekteedvqlmqi 179  
 QY 192 EDVFSQTLVDVNSLFRNSFNFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251  
 Db 180 envfsqtlvdvgylnmsfsvfkmgqgdflafqsfmsdtdsmeyffpafskapakka 239  
 QY 252 DLQOCWDIPNFFOLFNCNFSVIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311  
 Db 240 daepsalpnvfiqlcnlsfsvygsatvtemikaedlskqdsahggpsstcpvr 299  
 QY 312 DRGLCGELDONLSRCFKHEKQKQKQAHLSDECDPVPALHTEDEAIRLVNVSNOQYGOI 371  
 Db 300 drgsdglqgnlsdcvfnfrkrcqkcqylwadcpavpelytkadealelvnlsnqyqav 359  
 QY 372 LQMTREKHELTAYLVEKMRGQFCGWSELANOAPETETIIFNSIQVVPRI--HEGNISKODE 429  
 Db 360 lqmtqhlhdedtlylmekmregfgwvteiasqtpgseniifsklvvpgvhegnfskqdkm 419  
 QY 430 TMTDLSILPSSNFTLKIPLESSESSNFYGVYVAKALQHEKHEFTW 477  
 Db 420 tvvps-silpsnftlsspleksagnanfidhvvkvqlqhkhfkfw 466

RESULT 7

AAY30788

ID AAY30788 standard; Protein; 450 AA.

XX AAY30788;

XX AAY30788;

DT 23-NOV-1999 (first entry)

XX Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.

DE HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;

XX neuropsychiatric disorder; early-onset autosomal dominant myopia;

XX schizophrenia; splice variant.

[illegible]

Db	224	daeps	waipnfvllcnlsfsvyqsvsekittlratdpkqkdsnggppiskilpeq	283
Qy	312	DRGLCGELDQNLRCRCPFKHEKCKQCAHLSDEDCPDVPALHTELD	EAIRLVNSNQYGGI	371
Db	284	drsgdgklgulsdcvnrkrcdqylsdcdcpnvpelyrelnearlrvsrnqgdyq		343
Qy	372	LQWTRKHLEDTAVLVEKMRGQFGWSELANQAPETETIIFNSIOVVPRI	--HEGNISKQDE	429
Db	344	qvmqtqylndtllmekmrqfvgwslaygspgaedfnpvkvmvalsahegnssdqdd		403
Qy	430	TMWTDLSILPSSNFTKIPLEESAESSNFICYVVAKALQHFKEHFTW		477
Db	404	tvtps-sllpssnftlsspleksagnanfhdhvvkvlqhfkhehftw		450
RESULT	8			
AAV30789				
ID	AAV30789	standard; Protein; 374 AA.		
XX	AC	AAV30789;		
XX	DT	23-NOV-1999 (first entry)		
XX	DE	Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7c.		
XX	KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	KW	schizophrenia; splice variant.		
XX	OS	Cavia cobaya.		
XX	PN	WO9947535-A1.		
XX	PD	23-SEP-1999.		
XX	PF	16-MAR-1999; 99WO-US05606.		
XX	PR	16-MAR-1998; 98US-0078044.		
PR	PR	05-JUN-1998; 98US-0088312.		
PR	PR	28-OCT-1998; 98US-0106056.		
PR	PR	22-JAN-1999; 99US-0236134.		
XX	PA	(MILL-) MILLENNIUM PHARM INC.		
PA	PA	(REGC) UNIV CALIFORNIA.		
XX	PI	Chen H, Freimer NB;		
XX	DR	WPI: 1999-562047/47.		
DR	DR	N-PSDB; AAZ10757.		
XX	PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	PT	schizophrenia -		
XX	PS	Claim 1; Fig 9A-B; 205pp; English.		
XX	CC	The present sequence is encoded by a guinea pig HKNG1 ortholog		
CC	CC	gphkng1815 splice variant gphkng7c. HKNG1 (Hong Kong new gene 1)		
CC	CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	CC	polynucleotides are useful to identify compounds modulating HKNG1		
CC	CC	gene expression or HKNG1 polypeptide expression/activity. Compounds		
CC	CC	inhibiting or enhancing HKNG1 gene expression or activity in		
CC	CC	individuals can then be administered therapeutically to treat		
CC	CC	HKNG1-mediated disorders, especially neuropsychiatric disorders		
CC	CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	CC	in gene therapy techniques to treat such disorders. They are also used		
CC	CC	in diagnosis to identify individuals having, or at risk of developing		
CC	CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	CC	mutations especially result in the production of a protein with a		
CC	CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	CC	splice variant sequences, especially the substitution of a lysine for		

CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
CC useful in gene mapping, to produce probes or primers to identify  
CC similar sequences (e.g. mutants or sequences from different species)  
CC and to produce transgenic animals.  
XX  
SQ Sequence 374 AA;

Query Match 50.5%; Score 1277; DB 20; Length 374;  
Best Local Similarity 54.5%; Pred. No. 2,6e-99;  
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

QY 12 MKPPLLVFVCLLMKSHCAPTWKDKTAISENLKSFSEGEIDAEVKKALTIQKQK 71  
DB 1 mkpllmfpvcllwkdchcactwkdktaisenansfseageidvdgvekvialigkqm 60  
QY 72 IMMERKEHEHTNLMSTLKKCKREEKQKALKLNEVOHLEERLCRESLADSWGECRSL 131  
DB 61 immerreeehsklmktlkkckeeqkalklmevnehleeeslcqvsldswdecrac 120  
QY 132 ENNCMRIYTTQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191  
DB 121 esncmrfdttcqpawsvkn----- 140  
QY 192 EDVFSQLTVDVNSLNFNSFNFRMQQEPDQTFQSHFISDTDLTEPYFFPAFKEPMTKA 251  
DB 141 -----mepayra 147  
QY 252 DLEQCDIPNFQFCNFSVSIYESVETITKMLKATEDLPKDKAPDHGGLISKMLPGQ 311  
DB 148 daepswaipnvqllcnlsfsvyqsvsekittlratedppkqkdsnggppskilpeq 207  
QY 312 DRGLCGELDQNLSCRCFHEKCKQKQAHLSDECDPVPALHTELDIAIRLVNVSNOQYQGI 371  
DB 208 drgsdgklgnlsdcvnfrkrcqkqdylsddcpnvpeylrelnearlrvsrnqqydv 267  
QY 372 LQMTKRKHELDYALVEKMRGQFGWVSELANQAPETETIENSIOVVPRI--HEGNISKQDE 429  
DB 268 vqmtqghiedtllmekmrqfgwvsealayqspgaedifnpvkmvalsahegnssddqd 327  
QY 430 TMMTDLISLPSSNFTLKIPLEESAESSNFYGVYVAKALQHFKEHFTW 477  
DB 328 tvvps-sllpsnftlsspleksagnanfidhvvekvlgfhfkehtw 374

## RESULT 9

AA37090  
XX AAY30790 standard; Protein; 373 AA.  
AC AAY30790;  
XX 23-NOV-1999 (first entry)  
DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d.  
XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX Cavia cobaya.  
OS W09947535-A1.  
PN 23-SEP-1999.  
XX 16-MAR-1999; 99WO-US05606.  
XX 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX (MILL-) MILLENNIUM PHARM INC.

(REGC ) UNIV CALIFORNIA.

Chen H, Freimer NB;

WPI; 1999-562047/47.  
N-PSDB; AA210758.

New HKNG1 polynucleotides useful in diagnosis and treatment of  
neuropsychiatric disorders, e.g. bipolar affective disorders and  
schizophrenia -

Claim 1; Fig 10A-B; 205pp; English.

The present sequence is encoded by a guinea pig HKNG1 ortholog  
gphkng1815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)  
is a gene associated with bipolar affective disorder (BAD). HKNG1  
polynucleotides are useful to identify compounds modulating HKNG1  
gene expression or HKNG1 polypeptide expression/activity. Compounds  
inhibiting or enhancing HKNG1 gene expression or activity in  
individuals can then be administered therapeutically to treat  
HKNG1-mediated disorders, especially neuropsychiatric disorders  
e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
early-onset autosomal dominant myopia. The polynucleotides can be used  
in gene therapy techniques to treat such disorders. They are also useful  
in diagnosis to identify individuals having, or at risk of developing,  
HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
mutations especially result in the production of a protein with a  
different sequence to the human full-length HKNG1 polypeptide or  
splice variant sequences, especially the substitution of a lysine for  
a glutamic acid at residue 202 or 184. The polynucleotides are also  
useful in gene mapping, to produce probes or primers to identify  
similar sequences (e.g. mutants or sequences from different species)  
and to produce transgenic animals.

Sequence 373 AA;

Query Match 50.3%; Score 1272.5; DB 20; Length 373;  
Best Local Similarity 54.5%; Pred. No. 6.3e-99;  
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 12 MKPPLLVFVCLLMKSHCAPTWKDKTAISENLKSFSEGEIDAEVKKALTIQKQK 71  
DB 1 mkpllmfpvcllwkdchcactwkdktaisenansfseageidvdgvekvialigkqm 60

QY 72 IMMERKEHEHTNLMSTLKKCKREEKQKALKLNEVOHLEERLCRESLADSWGECRSL 131

DB 61 immerreeehsklmktlkkckeeqkalklmevnehleeeslcqvsldswdecrac 120

QY 132 ENNCMRIYTTQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191

DB 121 esncmrfdttcqpawsvkn----- 140

QY 192 EDVFSQLTVDVNSLNFNSFNFRMQQEPDQTFQSHFISDTDLTEPYFFPAFKEPMTKA 251

DB 141 -----mepayra 146

QY 252 DLEQCDIPNFQFCNFSVSIYESVETITKMLKATEDLPKDKAPDHGGLISKMLPGQ 311

DB 147 daepswaipnvqllcnlsfsvyqsvsekittlratedppkqkdsnggppskilpeq 206

QY 312 DRGLCGELDQNLSCRCFHEKCKQKQAHLSDECDPVPALHTELDIAIRLVNVSNOQYQGI 371

DB 207 drgsdgklgnlsdcvnfrkrcqkqdylsddcpnvpeylrelnearlrvsrnqqydv 266

QY 372 LQMTKRKHELDYALVEKMRGQFGWVSELANQAPETETIENSIOVVPRI--HEGNISKQDE 429

DB 267 vqmtqghiedtllmekmrqfgwvsealayqspgaedifnpvkmvalsahegnssddqd 326

QY 430 TMMTDLISLPSSNFTLKIPLEESAESSNFYGVYVAKALQHFKEHFTW 477

DB 327 tvvps-sllpsnftlsspleksagnanfidhvvekvlgfhfkehtw 373

```

RESULT 10
RAY30794
ID AAY30794 standard; Protein; 521 AA.
XX
AC AAY30794;
XX
DT 23-NOV-1999 (first entry)
XX
DE Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.
XX
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW schizophrenia; splice variant.
XX
OS Homo sapiens.
XX
PN WO9947535-A1.
XX
PD 23-SEP-1999.
XX
PF 16-MAR-1999; 99WO-US05606.
XX
PR 16-MAR-1998; 98US-0078044.
PR 05-JUN-1998; 98US-0088312.
PR 28-OCT-1998; 98US-0106056.
PR 22-JAN-1999; 99US-0236134.
XX
XX (MILL-) MILLENNIUM PHARM INC.
PA (REGC ) UNIV CALIFORNIA.
XX
PI Chen H, Freimer NB;
XX
DR WPI: 1999-562047/47.
DR N-PSDB; AAZ10762.
XX
PT New HKNG1 polynucleotides useful in diagnosis and treatment of
PT neuropsychiatric disorders, e.g. bipolar affective disorders and
PT schizophrenia.
XX
PS Claim 1: Fig 18A-B; 205pp; English.
XX
XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)
CC splice variant. HKNG1 is a gene associated with bipolar affective
CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds
CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/
CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or
CC activity in individuals can then be administered therapeutically to
CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as
CC early-onset autosomal dominant myopia. The polynucleotides can be used
CC in gene therapy techniques to treat such disorders. They are also useful
CC in diagnosis to identify individuals having, or at risk of developing,
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such
CC mutations especially result in the production of a protein with a
CC different sequence to the human full-length HKNG1 polypeptide or
CC splice variant sequences, especially the substitution of a lysine for
CC a glutamic acid at residue 202 or 184. The polynucleotides are also
CC useful in gene mapping, to produce probes or primers to identify
CC similar sequences (e.g. mutants or sequences from different species)
CC and to produce transgenic animals.
XX
SQ Sequence 521 AA;

Query Match 30.9%; Score 780.5; DB 20; Length 521;
Best Local Similarity 73.9%; Pred. No. 3.2e-57;
Matches 156; Conservative 8; Mismatches 16; Indels 31; Gaps 3;

QY 7 SNGSNKPPLLVIVICLLWLKDSHCAPTWKDKTAISENLKSPSEVGEIDADEVKALIG 66
:|||||
Db 118 nnsngmkppllvifvclllwikdshcptwkdktaisenlksfsevgdeidadeevkaltg 177

```

CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
CC sclerosis. In addition, (I) is involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth, and in tissue repair, healing of burns, incisions,  
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
CC disorders, or periodontal disease. Furthermore, (I) is also useful for  
CC gut protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, various immune deficiencies and  
CC disorders including severe combined immunodeficiency (SCID), bacterial or  
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
CC reactions and conditions, such as asthma or other respiratory problems.  
CC In addition, (I) affects biorhythms or circadian cycles of rhythms,  
CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
CC analgesic effects or other pain reducing effects, immunoglobulin like  
CC activity and can act as an antigen in a vaccine composition to raise an  
CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
CC amino acid sequences of the invention.  
XX  
SQ Sequence 449 AA;

Query Match 17.2%; Score 434; DB 22; Length 449;  
Best Local Similarity 25.8%; Pred. No. 4.3e-28;  
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;  
QY 12 MKPPLLVIVCLLWKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 65  
DB 1 mmktllllfvgl-----twesgqvlgdqtdsnelqemsgsgkyvknqiaavn 51  
QY 66 GIKQKIMMERKEKHTNLMSTLKKCRKQKALKLLNEVQHELEERLCRESLADSG 125  
DB 52 gvqkiltlektneerktllsneeakkkedalnetresetkikelpgvncetmmalwe 111  
QY 126 ECRSLNENCMRIYT-TCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEK 184  
DB 112 eckpcklqcmkfyarvcrgsglvgqlqleeflngspfyfwmngdr-----idsllen 165  
QY 185 DAQLTQMEDV----FSQLTVDVNSLNRSNFVPRMQQEQDQTFQSHFIS-DTDLTPEYF 239  
DB 166 drqqtthlmdvmqdhfrassiidelfqdrf-----ftrepqdy--hylpfslphrphf 218  
QY 240 FPAFSKEPMTKADLE-QCWDIPNFFOLFCNFSVSIYESVSETITKMLKAIEDLPKQDKAP 298  
DB 219 f--fksrivrslmpsfpyeplnfhmfqpflemlhea-----qqamdihfhsfaf 267  
QY 299 DHGGLSKMLPGQDGLGELDQNLRCFKFHEKCKQCOAHLSDEC-----PDVPALHTEL 354  
DB 268 qhpptefiregddrtvcvceirhnstgclrmkdqcdkreilsvdcstnnpqaklrrel 327  
QY 355 DEAIRLVNVSQOYGILQWTRKHELDYALVEKMRGQGWSELANQAPETEIFNSTQ 414  
DB 328 deslqvaeltrkynellksyqwkmlntsslleqlneqfnwvrsrlanltggedqyylyvt 387  
QY 415 VVPRIHGNIISKODETMMTDLSTLPSNFTLKIPLSESAESSNFIGYVYAKALQHF-KEH 473  
DB 388 tvas-htsdsvpsgvtevuvklfdspitvtvpvevsknbpkfmctvaekalqeyrkhh 446

RESULT 12  
AAR11704  
ID AAR11704 standard; Protein; 448 AA.  
XX  
AC AAR11704;  
XX

20-JUN-1991 (first entry)

Cytolysis Inhibitor.

KW cytolysis inhibitor; perforin; immunological effector molecule;  
KW infertility.  
XX Homo sapiens.  
XX Key Location/Qualifiers  
FH Peptide 1..21  
FT Protein /label= signal peptide  
FT Protein 22..226  
FT Protein /label= A-chain  
FT Protein 227..448  
FT Protein /label= B-chain  
XX DE3933850-A.  
PN 18-APR-1991.  
XX 06-OCT-1989; 89DE-3933850.  
XX 06-OCT-1989; 89DE-3933850.  
XX (SCHD ) SCHERING AG.  
XX Tschoopp J, Jenne D;  
XX WPI; 1991-118338/17.  
XX  
XX DNA sequence coding for cytolysis inhibitor - is strong inhibitor  
XX of terminal complement protein, eg perforin secreted by killer  
XX cells  
XX  
XX Claim 13; Page 9; 15pp; German.  
XX  
XX This cytolysis inhibitor is encoded by a 1.7kb BamHI-KpnI fragment  
XX isolated from a liver-specific cDNA library. It is a blood plasma  
XX component that inhibits immunological effector molecules. It is used  
XX for systemic or local treatment of inflammatory or autoimmune  
XX diseases mediated by complement or killer cells. It can also be  
XX used for detoxification of membrane-active and cytolytic proteins  
XX released by bacteria, fungi and insect venoms. The protein has a  
XX further use in the treatment of infertility caused by deficiency of  
XX cytolysis inhibitor. Monoclonal antibodies directed against the  
XX cytolysis inhibitor are also covered by the invention. They are used  
XX to isolate or quantify the natural protein in human plasma.  
XX See also AAQ11501 and AAQ11502.  
XX  
SQ Sequence 448 AA;

Query Match 17.1%; Score 432; DB 12; Length 448;  
Best Local Similarity 25.8%; Pred. No. 6.3e-28;  
Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 16 LLVFIVCLLWKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIKQ 69  
DB 4 lllfvgl-----twesgqvlgdqtdsnelqemsgsgkyvknqiaavnqvk 54  
QY 70 MKIMMERKEKHTNLMSTLKKCRKQKALKLLNEVQHELEERLCRESLADSGECRS 129  
DB 55 iktlektneerktllsneeakkkedalnetresetkikelpgvncetmmalweeekp 114  
QY 130 CLENENCMRIYT-TCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEKDAOL 188  
DB 115 clkqcmkfyarvcrgsglvgqlqleeflngspfyfwmngdr-----idsllenrdq 168  
QY 189 TQMEDV----FSQLTVDVNSLNRSNFVPRMQQEQDQTFQSHFIS-DTDLTPEYFFAF 243  
DB 169 thlmdvmqdhfrassiidelfqdrf-----ftrepqdy--hylpfslphrphff--f 219  
QY 244 SKEPMTKADLE-QCWDIPNFFOLFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHG 302  
DB 220 pksrivrslmpsfpyeplnfhmfqpflemlhea-----qqamdihfhsfafqhpp 270

Qy 303 LISKMLPGQDRGLGCELQNLISRCFKFHEKCKQKQAHLSDEC-----PDVPALHTLDEAI 358  
Db 271 tefiregddrtvcreifhstgclrmkdqcdkrellsvdcstnpsqaklrrldesi 330  
Qy 359 RLNVNSNOYQGIQLQMTKRKHLEDYALVEKMRGQGWSELANQAPETEIFNSIQVYPR 418  
Db 331 qvaerlrrkynellksyqwmntsslleqneqfnwvrlanltqgedgyllrvttvas 390  
Qy 419 IHEGNSIKQDQTMFDLSILPSSNFTLKIPLEESAESSNFYGYVAKALQHF-KEH 473  
Db 391 -htsdsrvpsgvtvkvkldsdplvtvpvsvrknkpfmetvaekalqeyrkkh 445

RESULT 13  
ABB50285  
ID ABB50285 standard; Protein: 416 AA.  
XX  
AC ABB50285;  
XX  
DT 08-FEB-2002 (first entry)  
XX  
DE Apolipoprotein J ovarian tumour marker protein, SEQ ID NO:60.  
XX  
KW Ovarian tumour marker gene; human; overexpression; upregulation;  
KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
KW identification; serous cystadenoma; borderline serous tumour;  
KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
KW immune response pathway; cell proliferation regulation; protein folding;  
KW membrane localised; secreted; therapeutic target; cytostatic;  
KW gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200175177-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 03-APR-2001; 2001WO-US10947.  
XX  
PR 03-APR-2000; 2000US-194336P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
XX  
DR WPI: 2001-626450/72.  
XX  
DR N-PSDB; ABA83111.  
XX  
PT Detecting and identifying ovarian tumor, identifying increased risk for  
PT developing ovarian cancer, and determining effectiveness of ovarian  
PT cancer treatment, by measuring expression level of ovarian tumor marker  
PT gene -  
XX  
PS Claim 23; Page 103-104; 140pp; English.  
XX  
XX The invention relates to methods for diagnosing and prognosing ovarian  
CC tumours in an individual via the detection and measurement of the  
CC expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,  
CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
CC ABA83181 and ABA83183). The methods of the invention are useful for  
CC detecting an ovarian tumour in a patient, for identifying an individual  
CC at increased risk for developing ovarian cancer, in prognostic tests for  
CC assessing the relative severity of ovarian cancer, in tests for  
CC monitoring a patient in remission from ovarian cancer and in tests for  
CC monitoring disease status in a patient being treated for ovarian cancer.  
CC The methods can additionally be used to identify a particular tumour as  
CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,

CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
CC tumour. The ovarian tumour marker genes of the invention were identified  
CC using SAGE (serial analysis of gene expression) and were found to be  
CC overexpressed in a broad variety of ovarian epithelial tumour cells  
CC relative to normal ovarian epithelial cells. The marker genes are  
CC implicated in immune response pathways, in the regulation of cell  
CC proliferation and in protein folding, and many of these are membrane-  
CC localised or secreted. In addition to their use as diagnostic and  
CC prognostic markers, the ovarian tumour marker genes or their encoded  
CC proteins may be used as therapeutic targets for the treatment and  
CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
CC proteins encoded by ovarian tumour marker genes of the invention.  
XX  
XX Sequence 416 AA;  
SQ

Query Match 16.08; Score 405.5; DB 22; Length 416;  
Best Local Similarity 25.98; Pred. No. 9.8e-26;  
Matches 113; Conservative 91; Mismatches 196; Indels 37; Gaps 12;  
Qy 49 SEVGIDADEVKKALTGIKQKIMMERKEHEHTNLMSTLKKCKREEKALKLLNEVOEH 108  
Db 2 snqskynkqinavngykiktlietnearktlisnleakkkedalnetresetk 61  
Qy 109 LEEERLCRESLADSWGECRSLENNCMRIYT-TCQPSWSSVKNKIERFFRKIYQFLPFP 167  
Db 62 lkeipgvneemmalweeckpckqtcmkfyarvcrsgslvgqrleeflnqsspfyfw 121  
Qy 168 HEDNEKDLPISEKLTLEKDAQLTQMEDV----FSQLTVDVNSLFNRSFNVRQMOQEPDOT 223  
Db 122 ngdr-----ldslendrqrqchmdvmqdhfrasslidelqgrf-----ftrepqdt 170  
Qy 224 FQSHFIS-DTDLTEPYFFPAFSKEPMTKADLE-QCWDIPNFFQFCNFSVSIYESVSETI 281  
Db 171 y--hylvfslphrrphff--fpkstrivrslnmpfspeplnfhmfqpfleimhea----- 221  
Qy 282 TKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLGCELQNLISRCFKFHEKCKQKQAHLS 341  
Db 222 ----ggamdihfshpafqhpptefiregddrtvcreirhnstgclrmkdqcdkrells 277  
Qy 342 EDC-----PDVPALHTLDEAIRLVNSNQYQGIQLQMTKRKHLEDYALVEKMRGQGWVS 397  
Db 278 vdcstnpsqaklrrldesiqvaerlrrkynellksyqwmntsslleqneqfnwvs 337  
Qy 398 ELANQAPETEIFNSIQVYPRHIEGNSIKQDQTMFDLSILPSSNFTLKIPLEESAESSN 457  
Db 338 rlanltggedgyllrvttvas-htsdsrvpsgvtvkvkldsdplvtvpvsvrknkpk 396  
Qy 458 FIGYVAKALQHF-KEH 473  
Db 397 fmetvaekalqeyrkkh 413

RESULT 14  
AAB03441  
ID AAB03441 standard; Protein: 446 AA.  
XX  
AC AAB03441;  
XX  
DT 03-JAN-2001 (first entry)  
XX  
DE Porcine clusterin protein sequence.  
XX  
KW Pig; clusterin; cell migration; wound healing; angiogenesis;  
KW cancer; vascular trauma; vascular disease; atherosclerosis;  
KW restenosis; complement cytotoxicity inhibitor; SP-40; 40; apoJ;  
KW testosterone repressed prostate message-2; sulfated glycoprotein-2.  
XX  
OS Sus scrofa.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /label= signal\_peptide





Search completed: July 2, 2002, 11:51:20  
Job time: 398 sec

	Novel human secreted proteins and nucleic acids for diagnosing, preventing and treating neurological, cardiovascular, infectious, autoimmune, gastrointestinal, bone disorders, cancer, particularly ovarian cancer
PT	
PT	
PT	
XX	
PS	Claim 11; Page 386-397; 421pp; English.
XX	
AAD08191-AAD08213	represent cDNAs corresponding to 19 human secreted protein genes and AAE03764-AAE03786 represent the proteins they encode. AAE03787-AAE03800 represent human secreted protein fragments or variants. The genes and their secreted proteins are useful for preventing, treating or ameliorating medical conditions, e.g., by protein or gene therapy. Pathological conditions can be diagnosed by determining the amount of the new protein in a sample or by determining the presence of mutations in the new genes. Specific uses are described for each of the 19 genes, based on the tissues in which they are most highly expressed, and include developing products for the diagnosis or treatment of proliferative disorders, cancer, tumours, foetal and developmental abnormalities, hematopoietic disorders, diseases of the immune system, AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation, allergies, neurological disorders (e.g., Alzheimer's disease, Parkinson's disease), cognitive disorders, schizophrenia, asthma, skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis, cardiovascular disorders, angiogenic disorders, kidney disorders, gastrointestinal disorders, pregnancy-related disorders, endocrine disorders, and infections. The proteins can also be used to aid wound healing and epithelial cell proliferation, to prevent skin aging due to sunburn, to maintain organs before transplantation, for supporting cell culture of primary tissues, to regenerate tissues, to identify their cognate ligands or binding partners, and in chemotaxis, and can be used as a food additive or preservative to modify storage properties. Antibodies specific for a protein of the invention can be used in alleviating symptoms associated with the disorders mentioned above, and in diagnostic immunoassays e.g., radioluminoassay or enzyme linked immunosorbent assay (ELISA). The present sequence represents a human secreted protein of the invention.
XX	
SQ	Sequence 363 AA;
Query Match	12.7%; Score 320; DB 22; Length 363;
Best Local Similarity	22.6%; Pred. No. 1.3e-18;
Matches	86; Conservative 80; Mismatches 136; Indels 88; Gaps 12;
QY	16 LLVFIIVCLLWLKDSHCAPTWKDKTAISLENLKSFSPVGIDAD-----EEVKALTCIQ 69
Db	:      ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   4 LIIlvglll -----twengrvigdmvstdeiqemstegskyinrelnaklvqvk 54
QY	70 MKTMEKEKEHTNLWMLTKCKREEQBALKLLNEVQHLEERLERLCRESIADSWGECRS 129
Db	:::  :   :   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   55 iktlieqtneerksllntleeakkkkedalndtkdsemklkasgvcvcdtmmlweeckp 114
QY	130 CLENNCMRIYT--TCOPSSSVNKNKERFRKIYQLFFPHEDNKDLPISEKLIEKAOL 188
Db	:  ::  ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   ::   115 clkctcmkfayrcvsrstglvhgvqeefingspfyfwngdr-----idsllendrqq 168
QY	189 TQMEDVFSQLTVDVNSLFNRSNFVRMQQEFDOTFQSHTSDLTPEFPFFAFSK--- 245
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 169 thaldv-----mqdsfdaasi-----mdelfqdrrft--raqdpfhfpssfq 213
QY	246 -----BPMYKADLECWDIPNFOLFPCMFVSIVSYSETITKM 284
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 214 rpfffnikhrfarnimpfgpyqpl-----nfdmfbgfjfmihqa-qgamdnv 260
QY	285 LKAIEDLPKDQAPDHGGGLISKMLPGDGRGLCGELDQNLSRCFHFHEKCCKCOAHLSEDC 344
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   :   : 261 lhrhpfpmftecdn-----qdgavcckearhnstgcclmmkqdcckecillsvdcc 310
QY	345 ----PDVPALHTELDFAIRL 360
Db	:   :   :   :   :   :   :   :   :   :   :   :   :   :   : 311 ssnnpaaqvlgcelnnslql 330

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:30 ; Search time 23.96 Seconds  
(without alignments)  
486.269 Million cell updates/sec

Title: US-09-722-544A-4MOD  
Perfect score: 2529  
Sequence: 1 MTRWYSNGNMKPLLVFI.....FTGVYVAKALQHKRHKFTW 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/6C\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/6D\_COMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2483	98.2	495	4	US-09-268-992-2
2	2460	97.3	466	4	US-09-268-992-64
3	2373	93.8	477	4	US-09-268-992-4
4	2346	92.8	446	4	US-09-268-992-51
5	1751.5	69.3	465	4	US-09-268-992-49
6	1635	64.7	466	4	US-09-268-992-39
7	1554	61.4	450	4	US-09-268-992-41
8	1277	50.5	374	4	US-09-268-992-43
9	1272.5	50.3	373	4	US-09-268-992-45
10	776	30.7	208	4	US-09-268-992-67
11	149	5.9	1388	2	US-08-685-576-1
12	145	5.7	1388	2	US-08-685-576-4
13	130.5	5.2	2482	1	US-08-328-254-6
14	130	5.1	3248	1	US-08-353-700-1
15	130	5.1	3248	5	PCT-US95-16216-1
16	126.5	5.0	1354	3	US-08-685-871-2
17	122	4.8	1886	4	US-08-938-105-3
18	121	4.8	828	2	US-08-993-228-21
19	120.5	4.8	1618	1	US-07-853-913-4
20	119.5	4.7	1786	4	US-08-973-462-8
21	119.5	4.7	3111	2	US-08-460-309-4
22	119.5	4.7	3111	2	US-08-125-077-4
23	118.5	4.7	816	2	US-08-533-306A-6
24	118.5	4.7	816	2	US-08-742-923A-6
25	117	4.6	1098	4	US-08-923-992A-8
26	117	4.6	1104	4	US-08-923-992A-4
27	117	4.6	1939	4	US-09-310-187A-1

28	117	4.6	2052	3	US-09-045-201A-2	Sequence 2, Appli
29	116.5	4.6	885	2	US-08-533-306A-4	Sequence 4, Appli
30	116.5	4.6	885	2	US-08-742-923A-4	Sequence 4, Appli
31	116.5	4.6	1057	4	US-09-541-782-10	Sequence 10, Appli
32	116	4.6	1164	4	US-08-923-992A-2	Sequence 2, Appli
33	114	4.5	435	2	US-08-531-439A-4	Sequence 4, Appli
34	114	4.5	1128	4	US-08-923-992A-6	Sequence 6, Appli
35	114	4.5	2154	2	US-08-841-349-4	Sequence 18, Appli
36	113.5	4.5	452	2	US-08-686-599A-18	Sequence 5, Appli
37	113.5	4.5	493	2	US-08-686-599A-5	Sequence 16, Appli
38	113.5	4.5	493	2	US-08-686-599A-16	Sequence 17, Appli
39	111.5	4.4	467	2	US-08-686-599A-17	Sequence 4, Appli
40	111.5	4.4	1066	4	US-09-541-782-8	Sequence 8, Appli
41	109	4.3	976	4	US-09-104-324B-4	Sequence 4, Appli
42	108	4.3	2285	4	US-09-308-375-2	Sequence 2, Appli
43	107.5	4.3	1038	4	US-09-541-782-4	Sequence 4, Appli
44	107	4.2	1164	4	US-08-923-992A-10	Sequence 10, Appli
45	107	4.2	1497	1	US-08-623-679-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-09-268-992-2  
; Sequence 2, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freilmer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 495  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-2

Query Match	98.2%;	Score	2483;	DB	4;	Length	495;
Best Local Similarity	99.6%;	Pred. No.	1.9e-230;				
Matches	469;	Conservative	2;	Mismatches	0;	Indels	0;
0;							
QY	7	NSNGNMKPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGETDADEEVKKALTG	66				
Db	25	NNSGNMKPLLVFIVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGETDADEEVKKALTG	84				
QY	67	IKQMKIMMERKEKHTNLMSTLKKCREKQKALKLLNEVQEHLEERLCRESLADSWGE	126				
Db	85	IKQMKIMMERKEKHTNLMSTLKKCREKQKALKLLNEVQEHLEERLCRESLADSWGE	144				
QY	127	CRSLENNCMRIYTCOPSSWSSVKNKIFERFRKIYQFLFPFHEDNEKDLPISEKLIKDA	186				
Db	145	CRSLENNCMRIYTCOPSSWSSVKNKIFERFRKIYQFLFPFHEDNEKDLPISEKLIKDA	204				
QY	187	QLTQMEDVFSQLTVDVNSLNFNRSENFVFRMQOEFDQTFQSHFISDTLTPEYFFPAFSKE	246				
Db	205	QLTQMEDVFSQLTVDVNSLNFNRSENFVFRMQOEFDQTFQSHFISDTLTPEYFFPAFSKE	264				
QY	247	PMTKADLEQCHWDINFFQLFCNFVSIVSVETITKMLKAIEDLPKODKAPDHGGLISK	306				

Db 265 PMTKADLEQCHDIPNFFQFCNFSVSIYESVETITKMLKAIEDLPQKDRAPDHGLISK 324  
Qy 307 MLPQDGLGCLGELDONLSRCFKFHEKQKCOAHLSDECDDVPALHTELDEAIRLVNSNQ 366  
Db 325 MLPQDGLGCLGELDONLSRCFKFHEKQKCOAHLSDECDDVPALHTELDEAIRLVNSNQ 384  
Qy 367 QYQTLQWTRKHLEDATYLVKMRGQFCWSELANOAPETETIIFNSIQVVPRIHEGNISK 426  
Db 385 QYQTLQWTRKHLEDATYLVKMRGQFCWSELANOAPETETIIFNSIQVVPRIHEGNISK 444  
Qy 427 QDETMTDLSILPSSNFTLKIPLESASSNFVIGYVAKALQHKEHFKTW 477  
Db 445 QDETMTDLSILPSSNFTLKIPLESASSNFVIGYVAKALQHKEHFKTW 495

RESULT 2  
US-09-268-992-64  
; Sequence 64, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 64  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-64

Query Match 97.3%; Score 2460; DB 4; Length 466;  
Best Local Similarity 99.8%; Pred. No. 2.8e-228;  
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
Qy 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEYKALTGIKQMK 71  
Db 1 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEYKALTGIKQMK 60  
Qy 72 IMMERKEHTNLMTLKKCREKQKALLLNEVQHELEERLCRESLADSGECSRL 131  
Db 61 IMMERKEHTNLMTLKKCREKQKALLLNEVQHELEERLCRESLADSGECSRL 120  
Qy 132 ENNCMRIYTTCPQSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191  
Db 121 ENNCMRIYTTCPQSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 180  
Qy 192 EDVFSQTLVDVNSLNFNSFNVMQOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251  
Db 181 EDVFSQTLVDVNSLNFNSFNVMQOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 240  
Qy 252 DLEQCDWIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 311  
Db 241 DLEQCDWIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 300  
Qy 312 DRGLCGELDONLSRCFKFHEKQKCOAHLSDECDDVPALHTELDEAIRLVNSNQYQGI 371  
Db 301 DRGLCGELDONLSRCFKFHEKQKCOAHLSDECDDVPALHTELDEAIRLVNSNQYQGI 360

Qy 372 LQWTRKHLEDATYLVKMRGQFCWSELANOAPETETIIFNSIQVVPRIHEGNISKODETM 431  
Db 361 LQWTRKHLEDATYLVKMRGQFCWSELANOAPETETIIFNSIQVVPRIHEGNISKODETM 420  
Qy 432 MTDLSILPSSNFTLKIPLESASSNFVIGYVAKALQHKEHFKTW 477  
Db 421 MTDLSILPSSNFTLKIPLESASSNFVIGYVAKALQHKEHFKTW 466

RESULT 3  
US-09-268-992-4  
; Sequence 4, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 477  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-4

Query Match 93.8%; Score 2373; DB 4; Length 477;  
Best Local Similarity 91.3%; Pred. No. 6.7e-220;  
Matches 454; Conservative 2; Mismatches 1; Indels 40; Gaps 2;  
Qy 1 MRTWDSNSGNMKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEV 60  
Db 1 MRTWDSNSGNMKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEEV 60  
Qy 61 KKALTGIKQKIMMERKEK-----EHTNLMSTLKKCREKQKQALK 100  
Db 61 KKALTGIKQKIMMERKEKANOAPETETIIFNSIQVVPRIEHTNLMSTLKKCREKQKQALK 120  
Qy 101 LLNEVQHELEERLCRESLADSGECSRLNENCMRIYTTCPQSWSSVKNKIERFRKI 160  
Db 121 LLNEVQHELEERLCRESLADSGECSRLNENCMRIYTTCPQSWSSVKNKIERFRKI 180  
Qy 161 YQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQTLVDVNSLNFNSFNVMQOE 220  
Db 181 YQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQTLVDVNSLNFNSFNVMQOE 240  
Qy 221 DQTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEOCWDIPNFFQFCNFSVSIYESV 280  
Db 241 DQTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEOCWDIPNFFQFCNFSVSIYESV 300  
Qy 281 ITKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQKCOAHL 340  
Db 301 ITKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQKCOAHL 360  
Qy 341 SEDCDVPALHTELDEAIRLVNSNQYQGIQWTRKHLEDATYLVKMRGQFCWSEL 400  
Db 361 SEDCDVPALHTELDEAIRLVNSNQYQGIQWTRKHLEDATYLVKMRGQFCWSEL- 419  
Qy 401 NOAPETETIIFNSIQVVPRIHEGNISKODETMTDLSILPSSNFTLKIPLESASSNF 460  
Db 420 -----HEGNISKODETMTDLSILPSSNFTLKIPLESASSNF 460

QY 461 YVAKALQHKEHFKTW 477  
|||||  
Db 461 YVAKALQHKEHFKTW 477

## RESULT 4

US-09-268-992-51  
; Sequence 51, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 51  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-51

Query Match 92.8%; Score 2346; DB 4; Length 446;  
Best Local Similarity 99.8%; Pred. No. 2.4e-217;  
Matches 445; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 32 APTWKKTATSENKSFSEVGEIDAEVKKALTGKIMMERKEKEHTNLMSTLKKC 91  
|||||  
Db 1 APTWKKTATSENKSFSEVGEIDAEVKKALTGKIMMERKEKEHTNLMSTLKKC 60  
QY 92 REKQKALKNLNEVOHLEERLCRESLADSWGECRSCLNENCMRIYTCOPSSSVKN 151  
|||||  
Db 61 REKQKALKNLNEVOHLEERLCRESLADSWGECRSCLNENCMRIYTCOPSSSVKN 120  
QY 152 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOLTQMEDVFSOLTVDVNSLNRSEN 211  
|||||  
Db 121 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOLTQMEDVFSOLTVDVNSLNRSEN 180  
QY 212 VFQMQQEFQDTQSHFISDTLTETPYFPFAPSKPMTRADLEQCQWDIPNFFQLCFNFSV 271  
Db 181 VFQMQQEFQDTQSHFISDTLTETPYFPFAPSKPMTRADLEQCQWDIPNFFQLCFNFSV 240  
QY 272 SIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGDQKGLCGELDONLSRCFKPHE 331  
Db 241 SIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGDQKGLCGELDONLSRCFKPHE 300  
QY 332 KKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYGOILOMKRKHLEDYALVEKMRG 391  
Db 301 KKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYGOILOMKRKHLEDYALVEKMRG 360  
QY 392 QFGWSELANQAPETIIINFSQVVPRIHEGNIKSQDETMVTDLSILPSSNFTLKIPLKE 451  
Db 361 QFGWSELANQAPETIIINFSQVVPRIHEGNIKSQDETMVTDLSILPSSNFTLKIPLKE 420  
QY 452 SAESSNFIYVAKALQHKEHFKTW 477  
|||||  
Db 421 SAESSNFIYVAKALQHKEHFKTW 446

## RESULT 5

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

US-09-268-992-49  
; Sequence 49, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 49  
; LENGTH: 465  
; TYPE: PRT  
; ORGANISM: Bos sp.  
US-09-268-992-49

Query Match 69.3%; Score 1751.5; DB 4; Length 465;  
Best Local Similarity 70.6%; Pred. No. 4e-160;  
Matches 329; Conservative 59; Mismatches 77; Indels 1; Gaps 1;  
QY 12 MKPPLLVIYVCLLNLKDSHCAPTWDKTAISENLKSFSEVGEIDAEVKKALTGKIMK 71  
Db 1 MKPPLLVIYVCLLNLKDSHCAPTWDKTAISENLKSFSEVGEIDAEVKKALTGKIMK 60  
QY 72 IMMERKEHTNLMSTLKKCREKEQKALKLNEVOHLEERLCRESLADSWGECRSCL 131  
Db 61 IMMERKEHTNLMSTLKKCREKEQKALKLNEVOHLEERLCRESLADSWGECRSCL 120  
QY 132 ENNCRIYTCOPSSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOLTQ 191  
Db 121 ESDCMRFYTCOSSWSMKSTIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAOLTQ 180  
QY 192 EDVFSQLTVDVNSLNRSENFMFROMQOEFDQTFQSHFISDTLTETPYFPFAPSKPMTRKA 251  
Db 181 ENVFSQLTVDVNSLNRSENFMFROMQOEFDQTFQSHFISDTLTETPYFPFAPSKPMTRKA 240  
QY 252 DLEQCDIPNFFQLCFNFSVSIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGO 311  
Db 241 HPMQSWDIPNFFQLCFNFSVSIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGO 300  
QY 312 DRGLCGELDONLSRCFKPHEKQCAHLSEDCPDVPALHTELDEAIRLVNVSNOQYGOI 371  
Db 301 GRGLCGEPGQNSSECLQFHARCKQCDYLWADCPAVPELYTKADEALVELVNSNOQYAOV 360  
QY 372 LQMTKRKHLEDYALVEKMRGQFGWSELANQAPETIIINFSQVVPRIHEGNIKSQDETM 431  
Db 361 LQMTKRKHLEDYALVEKMRGQFGWSELANQAPETIIINFSQVVPRIHEGNIKSQDETM 420  
QY 432 MTDLSILPSSNFTLKIPLKEESAESSNFIYVAKALQHKEHFKTW 477  
Db 421 I-DISILPSSNFTLTIPLKEESAESSNFIYVAKALQHKEHFKSW 465

## RESULT 6

US-09-268-992-39  
; Sequence 39, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING



; SEQ ID NO 43  
; LENGTH: 374  
; TYPE: PRT  
; ORGANISM: Cavia sp.  
US-09-268-992-43

Query Match 50.5%; Score 1277; DB 4; Length 374;  
Best Local Similarity 54.5%; Pred. No. 1.3e-114;  
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;  
QY 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTGKQMK 71  
Db 1 MKPLLMFPVCLLWLDKSHCAPTWKDKTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60  
QY 72 IMMERKEHTNLMSTLKKREEKQKALKNVEQHEERLCRESLADSWGECRSCL 131  
Db 61 IMMERREEHSLKMTLKKCKEKEQKALKNVEVHEHEERESLCQVSLADSWDECRACL 120  
QY 132 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191  
Db 121 ENSCMRFTTCCQPAWSSVKN----- 140  
QY 192 EDVFSQLTVDVNSLFRNSFNVRMQQEFDDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251  
Db 141 -----MEPAYRA 147  
QY 252 DLEQCWDIPNFPQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 311  
Db 148 DAEPSSWAIPNVFOLLNLSFSVQSVSEKLTTLRATEDPPKQKDSNOGGPISKILPEQ 207  
QY 312 DRGLGELDONLSRCFHEKCKQKQAHLSDCDPPVPAHTELDDEAIRLVNVSNOQYGOI 371  
Db 208 DRGSDGLQGNLSDCVNFRKRCQKQDYLSDDCPNVPYELRELNEALRLVSRNSQYDQV 267  
QY 372 LQMTKRKHELTAYLVEKMRGQFGWVSELANOAPETIEIENSIOVVPRI--HEGNISKODE 429  
Db 268 VQMTQYHLEDTTLLMEKMRQFGWVSELAYQSPGAEDIFNPVKVVALSAHSGNSDQDD 327  
QY 430 TMMTDLSILPSSNFTLKIPLESASSNFIYGVYVAKALQHFKEHFKTW 477  
Db 328 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 374

RESULT 9  
US-09-268-992-45  
; Sequence 45, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 45  
; LENGTH: 373  
; TYPE: PRT  
; ORGANISM: Cavia sp.  
US-09-268-992-45

Query Match 50.3%; Score 1272.5; DB 4; Length 373;  
Best Local Similarity 54.5%; Pred. No. 3.4e-114;  
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;  
QY 12 MKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTGKQMK 71  
Db 1 MKPLLMFPVCLLWLDKSHCAPTWKDKTAISENANFSFSEAGEIDVDGEVKIALIGIKQMK 60  
QY 72 IMMERKEHTNLMSTLKKREEKQKALKNVEQHEERLCRESLADSWGECRSCL 131  
Db 61 IMMERREEHSLKMTLKKCKEKEQKALKNVEVHEHEERESLCQVSLADSWDECRACL 120  
QY 132 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQM 191  
Db 121 ENSCMRFTTCCQPAWSSVKN----- 140  
QY 192 EDVFSQLTVDVNSLFRNSFNVRMQQEFDDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 251  
Db 141 -----MPAY-----RA 146  
QY 252 DLEQCWDIPNFPQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 311  
Db 147 DAEPSSWAIPNVFOLLNLSFSVQSVSEKLTTLRATEDPPKQKDSNOGGPISKILPEQ 206  
QY 312 DRGLGELDONLSRCFHEKCKQKQAHLSDCDPPVPAHTELDDEAIRLVNVSNOQYGOI 371  
Db 207 DRGSDGLQGNLSDCVNFRKRCQKQDYLSDDCPNVPYELRELNEALRLVSRNSQYDQV 266  
QY 372 LQMTKRKHELTAYLVEKMRGQFGWVSELANOAPETIEIENSIOVVPRI--HEGNISKODE 429  
Db 267 VQMTQYHLEDTTLLMEKMRQFGWVSELAYQSPGAEDIFNPVKVVALSAHSGNSDQDD 326  
QY 430 TMMTDLSILPSSNFTLKIPLESASSNFIYGVYVAKALQHFKEHFKTW 477  
Db 327 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 373

RESULT 10  
US-09-268-992-67  
; Sequence 67, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 67  
; LENGTH: 208  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-67

Query Match 30.7%; Score 776; DB 4; Length 208;  
Best Local Similarity 98.6%; Pred. No. 8.1e-67;  
Matches 145; Conservative 2; Mismatches 0; Indels 0; Gaps 0;  
QY 7 SNSGNMKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 66  
Db 57 NNSGNMKPPLLVFVCLLWLDKSHCAPTWKDKTAISENLKSFSEVGEIDAEVVKALTG 116

QY 67 IKQKIMMERKEHTNLMSTLKKCREKQKALKLLNEVQEHLEEBERLCRESLADSWGE 126  
DB 117 IKQKIMMERKEHTNLMSTLKKCREKQKALKLLNEVQEHLEEBERLCRESLADSWGE 176

QY 127 CRSCLENNCMRIYTCQPSWSVKNKI 153  
DB 177 CRSCLENNCMRIYTCQPSWSVKNKL 203

RESULT 11  
US-08-685-576-1  
; Sequence 1, Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaibuchi, Koza  
; APPLICANT: Iwamatsu, Akihiro  
; APPLICANT: Nakano, Takeshi  
; APPLICANT: Ito, Masaaki  
; APPLICANT: Takahashi, No. 5906819uaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,576  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-325129  
; FILING DATE: 20-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-17150  
; FILING DATE: 05-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-131206  
; FILING DATE: 26-APR-1996  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16887/843  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1388 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-08-685-576-1

Query Match 5.9%; Score 149; DB 2; Length 1388;  
Best Local Similarity 16.7%; Pred. No. 3e-05;  
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISENLKSFSEVEIDEADEVKALTGKQMKI-----MMERKEKEHTNLM 86  
DB 780 KQKDVNLNEDVRLNLT-----LKIEQETQKRCULTQNDLRKMTQOVNLTLMKSEKQKQENHLL 835  
QY 87 TLK-----KCREKQKALKLLNEVQEHLEEBE-----RLCR 117

DB 836 EMKMSLEKQNAELRKERQDADGQMKELQDLQAEAEQYFTSTLYKTQVRELKEECEKTKLCK 895  
QY 118 E-----SLADSGEGRSCLENNCMRIYTCQPSWSVKNKIERRFKIYQFLFPFHEDN 171  
DB 896 ELQOKKQELQDE-----RDSLAQALEITLTAKDSQALARSIAEEQYSDLEKE 942  
QY 172 E--KDLPISE-----KLIEKDAQLTQMBEDVFSQLTVDVNSLFRNSFNVRMQQEFQDQ 222  
DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLTSDVANLANEKEELNNKLKEAQEQ 1002  
QY 223 TFQSHFISDLDLTPYFPFAPFSKPEMTKADLEQCDWDIPNEFFQLFCNFSVSIYESVETIT 282  
DB 1003 LSR---LKDEEISAAATKAQFEXQLLTERLK-----TQAVN 1036  
QY 283 KMLKAIE-----DLPKQDKAPDGHGLISKMLPGQDRGLCGEL-----DQNLISRCF 327  
DB 1037 KLAEMNRKEPVKRGNDTDVRRKEK-----ENRKLHMLKSEREKLTOQMI 1082  
QY 328 KFHEKQKQAHLSDECDPVPALHTELDEAIRLVNVSNNQYQGILQWTRKHLE-----380  
DB 1083 KYQKELNEMOQAIAEE---SQIRIELQMTLDSKSDIEQLRSQQLQALHIGLOSSSTGSG 1138  
QY 381 -----DTAYLVEKMRG-----QEGWVSEL-----399  
DB 1139 PGDTEADDDGFPESRLEGLWSLPLVRNNTKKGGWVKYVIVSVSKKILFYDSEQDKESQSNPYM 1198  
QY 400 -----ANQAPETETIENSIOQVPRI-----HEGNISKQDETMMWTLDSILPSSNFT 444  
DB 1199 VLDIDKLFHVRPVQTQDVYRADAKEIPRIFOILYANEGESKKEQE-----1243  
QY 445 LKIPLEESAESSNFI---GYVVAKALQHF 470  
DB 1244 --FPVEPVGEKSNVICHKGHEFIPTLYHF 1270

RESULT 12  
US-08-685-576-4  
; Sequence 4, Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaibuchi, Koza  
; APPLICANT: Iwamatsu, Akihiro  
; APPLICANT: Nakano, Takeshi  
; APPLICANT: Ito, Masaaki  
; APPLICANT: Takahashi, No. 5906819uaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,576  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-325129  
; FILING DATE: 20-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-17150  
; FILING DATE: 05-JAN-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-131206  
; FILING DATE: 26-APR-1996



ATTORNEY/AGENT INFORMATION:  
NAME: Bent, Stephen A.  
REGISTRATION NUMBER: 29,768  
REFERENCE/DOCKET NUMBER: 16887/843  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202)672-5300  
TELEFAX: (202)672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1388 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-685-576-4

Query Match 5.7%; Score 145; DB 2; Length 1388;  
Best Local Similarity 17.0%; Pred. No. 7.3e-05;  
Matches 95; Conservative 96; Mismatches 176; Indels 192; Gaps 20;  
QY 36 KKTALSENKSFSEVGEIDAEVYKALTGKIKOMKI-----MMERKEKEHTNLS 86  
DB 780 KQKVDLNEVRLT-----LKIEQTKRCLTQNDLKMTQVNTLKSEKOLKQENHLM 835  
QY 87 TLK-----KCREKQKALKNVQPHLEEE-----RLCRESLADSWGEC----- 127  
DB 836 ENKMLERQNALRERQDQOMKELQDLQEAQYFSTLYKTQVRELKECEKTKLKG 895  
QY 128 ---RCLNENCMRYITTCQPSWSSVKNKIERFFRYQFLFPFHEDNE--KOLPIS-- 179  
DB 896 ELQKKQELQDRSLAAQLTTLTKADSEQLARSIAEQYSDLEKIMKEIKEMMA 955  
QY 180 ---KLIKDAQLTQMEDVFSQLVNLSFNRS---FNVRQMQQEDQTFQSHFISDT 232  
DB 956 RHQBELTEKDIATISLEETNTLSDVANLANEKEELNKLKDVQELSR-----LKDE 1009  
QY 233 DLTEPYFPFAPSKPEMTKADLEQCDWIPNFFQPCNFSVSVIESVETITKMLKAE-- 289  
DB 1010 EISAAIKAQPEKQLTERTLK-----TQAVNKLAEIMNRKE 1046  
QY 290 -----DLPKQKADPHGLISKMLPGQDRGLCGEL-----DONLSRCFKFHEKQKQ 337  
DB 1047 PVKRGNDTDVRRKE-----ENRKLHMLKSEREKLQOMIKYQKELNEMQ 1092  
QY 338 AHLSEDCPDVPAHTELDEAIRLVNVSNOQYQQLQWTRKHL-----DTAY 384  
DB 1093 AQIAEE-----SQRIELQWTLDSKSDIEQLRSQALHGLDSSIGSGPGDAEADDG 1148  
QY 385 LVEKWRG-----QFGWVSEL-----A 400  
DB 1149 PESLEGWLSLVRRNNTKFGVKKYVIVSSKKILFYDSEQKQSNPYWVLIDIKLHV 1208  
QY 401 NQAPETEIFNSIQVVPRI-----HEGNISKQDETMTDLSLPSNFTLKIPLEESAE 454  
DB 1209 RPTQTDVYRADAKEIPRIFQILYANEGESKKEQ-----FPVEPVGE 1251  
QY 455 SSNFI---GYVAKALQHF 470  
DB 1252 KSNYICHKGHEFTPLYHF 1270

RESULT 13  
US-08-328-254-6  
Sequence 6, Application US/08328254  
Patent No. 5710022  
GENERAL INFORMATION:  
APPLICANT: Zhu, Xuelling  
APPLICANT: Lee, Wen-Hwa  
TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700  
CITY: San Diego  
STATE: California  
COUNTRY: USA  
ZIP: 92122  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/328,254  
FILING DATE: 24-OCT-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/141,239  
FILING DATE: 22-OCT-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Campbell, Cathryn A.  
REGISTRATION NUMBER: 31,815  
REFERENCE/DOCKET NUMBER: P-CJ 1191  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (619) 535-9001  
TELEFAX: (619) 535-8949  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 2482 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 5.2%; Score 130.5; DB 1; Length 2482;  
Best Local Similarity 18.8%; Pred. No. 0.0044;  
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;  
QY 27 KDSHCAPTWKDKTALSENK-----SFSEVGE-----IDADE---EYKALTGIQ 69  
DB 1502 KOLHIAEKLKERENDSLKQVENLERELQMSQENQELVILDAENSKAEVETLKTQIEE 1561  
QY 70 M-----KIMERKEKEH-----TNLMSTLKKCKREEKQKALKNLNEV 105  
DB 1562 MARSLKVFELDLVTLRSEKENLTQIQEQKQSELQKLLSSFKSLLEKEQA-----EI 1616  
QY 106 OEHLEERELCRSLADSWGECRSCLENNC--MRIYTCQPSWS-----SVKNKIERF 156  
DB 1617 Q--IKESKTAVELQNLKELNEVAALCGDQEIWKATEQSLDPPTEEEHQLRNSIEKL 1674  
QY 157 FRKIYQFLFPFHEDNEKDLPISEKLEKD-----185  
DB 1675 RARL-----EADKKQLCVLQOLKESEHDLKGRVENLERELEIARTNOEHALEA 1727  
QY 186 -----AQLTQMEDVFSQTLVDVNSLFNRSNVFRMQQEDQTFQSHFISDTDL 234  
DB 1728 ENSGVEVETLKAKTEGTSRLGLELDVVTIRSEKENLTNELQKEQER-----ISELEI 1781  
QY 235 TEPYFFPAPSKPEMTKADLEQCDWIPNFFQPCNFSVSVIESVETITKMLK-IEDLPK 293  
DB 1782 INSSFENILQKEQKQVQKME-----KSSSTAMEMLOTQLKELNE 1820  
QY 294 QDKAPDHGGLISKMLPGQDRGLCGELQDNLS---RCFKFHEKQKQAHLSQEDCPDVPAL 350  
DB 1821 RVAALHN-----DQEAACKAKEQLNSQVECLEL-EKAQLLQ-----1856  
QY 351 HTEDEAIRLVNVSNOQYQQLQWTRKHLDETALVLEKMRGQFGWVSELANQAPETEIF 410  
DB 1857 ---LDEAKNNYIVLQSSVNGLIQ-----EVEDQKQLKEDDE---ISRLKNOIQDEQIV 1906  
QY 411 NSIQVVPRIH-----EGNISKQDETMTD--LSLPSNFTLKIPLE 450  
DB 1907 SKLSQVEGEHQLWKQONLELRNLTVLEQKIQVLOSKNASLQDTLE 1952

Db 2448 QEHAALAEANSKEVETLKAIEGTQSLRGLDLDVVTIRSEKENTNELQKEQERISEL 2507  
QY 328 -----KPEKCO-----KQAHLSDCPDVPALHTE-----L 354  
Db 2508 EIINSFENILQEKQVKQMEKSKSTAMEMLOTOLKELNERVAALRNDQACKAKEQNL 2567  
QY 355 DEAIRLVNVSNOQYGO-----ILQMRKHL-----EDTAYLVEKMRQFGWVSELA 400  
Db 2568 SSOVECLELEKQALQGLDEAKNNYIVLOSSYKGLIQEVEDGKQKLEKDEE---ISRLK 2624  
QY 401 NOAPETETIIFNSIQVVPRIH-----EGNISKQDETMTD---LSILPSSNFTLKIPLE 450  
Db 2625 NOIQDQEQVLVSKLSQVEGEHQLWKQENLRNLTVLEQKIQVLOSKNASLODTLE 2680

RESULT 15  
PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; APPLICANT: Ratner, Jerome B.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochores Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16216  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/353,700  
; FILING DATE: 09-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36,252  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; PCT-US95-16216-1

Query Match 5.1%; Score 130; DB 5; Length 3248;  
Best Local Similarity 19.6%; Pred. No. 0.0074;  
Matches 105; Conservative 94; Mismatches 179; Indels 158; Gaps 28;

QY 37 DKTAISENLK-SFSEVGEI-----DADEVKALTGKIKIMMERKEKHTNLMSTLKK 90  
Db 2181 EKLNVSKALEAALVEKGEFALRLSTQEEVHQRRGIEKLVRVRIEADKKQLHIAEKLKE 2240  
QY 91 CREKQEQALKLLNEVQHELEERLCRES-----LADSWGECRSCLENNCMRYITTCQPSW 146  
Db 2241 -RERENDSLK--DKVENLERELQMSQENQELVILDA-----ENSKAEVET----- 2282

RESULT 14  
US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; APPLICANT: RATNER, JEROME B.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
; US-08-353-700-1

Query Match 5.1%; Score 130; DB 1; Length 3248;  
Best Local Similarity 19.6%; Pred. No. 0.0074;  
Matches 105; Conservative 94; Mismatches 179; Indels 158; Gaps 28;

QY 37 DKTAISENLK-SFSEVGEI-----DADEVKALTGKIKIMMERKEKHTNLMSTLKK 90  
Db 2181 EKLNVSKALEAALVEKGEFALRLSTQEEVHQRRGIEKLVRVRIEADKKQLHIAEKLKE 2240  
QY 91 CREKQEQALKLLNEVQHELEERLCRES-----LADSWGECRSCLENNCMRYITTCQPSW 146  
Db 2241 -RERENDSLK--DKVENLERELQMSQENQELVILDA-----ENSKAEVET----- 2282

QY 147 SSVKNKIERFFR--KIYQ-LFPFHEDNEKOLPISEKLIKEDAOITQMEDVFS----- 196  
Db 2283 --LKTQIEEMARSLKIFELDLVTLRSEKEN---LTKQIEKQEQQLSELDKLSSFKSLLE 2337  
QY 197 ---QLTVDVNSLFNSFNFRQMOEQFQTFQSHFISDTDL--TEPYFFPAFSKEPMTK 250  
Db 2338 EKEQAEIQKESKTAVMQLNQKELNEAVAA-LCGQIEKMKATEQSLDPPIEEHQLR 2396  
QY 251 ADLEQCDIPNPFQCFNFSVIYSV-----SEITIMKLA-IEDLPKQDKA----- 297  
Db 2397 NSIE-----KLRLAEADKKQLCVLOQLKESEHADLLKGRVENLERELEIARTN 2447  
QY 298 PDHGGI-----ISKMLPGQDRGLCG-ELD-----QNLRCF----- 327

Search completed: July 2, 2002, 11:52:33  
Job time: 216 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:00 ; Search time 33.04 Seconds

(without alignments)  
1387.246 Million cell updates/sec

Title: US-09-722-544A-4MOD

Perfect score: 2529

Sequence: 1 MRTWDYSNGNMKRPPLLVFL.....FIGYVAKALQHFKEHFXTW 477

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	% Query Match	Length DB	ID	Description
1	434	17.2	449	1 A41386	clusterin precursor
2	418	16.5	451	1 I50131	clusterin - quail
3	410.5	16.2	445	2 A40018	clusterin precursor
4	408	16.1	449	2 S07714	T64 protein precursor
5	403	15.9	448	2 A40714	clusterin precursor
6	402	15.9	448	2 I56335	apolipoprotein J -
7	400.5	15.8	439	2 A35744	clusterin precursor
8	400.5	15.8	446	2 A42108	clusterin precursor
9	380	15.0	447	2 A27205	clusterin precursor
10	168	6.6	191	2 I48174	sulfated glycoprot
11	164	6.5	1738	2 T14867	interaplin - slime
12	151	5.9	11526	2 T41522	myosin ii - fission
13	150	5.9	11356	2 B70356	chromosome assembl
14	149	5.9	1388	2 S70633	serine/threonine-s
15	148.5	5.9	1091	2 T34107	hypothetical prote
16	147	5.8	3225	2 I52300	giantin - human
17	147	5.8	3259	1 A56539	giantin - human
18	145.5	5.8	594	1 A44073	C1K1 protein - yea
19	141.5	5.6	1005	2 A64465	hypothetical prote
20	141.5	5.6	3433	1 S28381	utrophin - human
21	140.5	5.6	2166	2 G70163	hypothetical prote
22	140	5.5	1132	2 T00259	hypothetical prote
23	139.5	5.5	2863	1 S28261	centromere protein
24	138	5.5	1084	2 G71329	hypothetical prote
25	137	5.5	1955	2 T30934	myosin heavy chain
26	137	5.4	1940	1 S04090	myosin heavy chain
27	136	5.4	1388	2 S74245	serine/threonine-s
28	135.5	5.4	1679	2 S48385	hypothetical prote
29	135.5	5.4	1956	2 T16416	hypothetical prote

30	135	5.3	1060	1 A40264	kinesin-related pr
31	135	5.3	1133	2 T22976	hypothetical prote
32	134.5	5.3	1300	2 I53799	Cgl protein - huma
33	134	5.3	1538	2 T28095	cardiac muscle fac
34	133.5	5.3	1313	2 A48467	myosin heavy chain
35	133.5	5.3	1410	1 A57013	early endosome ant
36	133.5	5.3	1957	2 T38077	hypothetical colle
37	132.5	5.2	852	2 D72230	conserved hypothet
38	132	5.2	1033	2 T42701	hypothetical prote
39	132	5.2	1992	2 A47297	myosin heavy chain
40	131.5	5.2	1979	1 S03166	myosin heavy chain
41	131	5.2	1024	2 T34517	kinesin-related pr
42	131	5.2	1413	2 T26467	hypothetical prote
43	130.5	5.2	1940	1 A24922	myosin heavy chain
44	130	5.1	746	2 T21277	hypothetical prote
45	130	5.1	793	2 S34830	kinesin-related pr

#### ALIGNMENTS

RESULT 1  
A41386  
clusterin precursor [validated] - human  
N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein  
N:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Homo sapiens (man)  
C:Date: 03-Apr-1992 #sequence revision 17-Nov-1995 #text-change 08-Dec-2000  
C:Accession: S43646; S04662; A41386; A35833; S34056; A5177; A37816; B37816; PL0136;  
R:Wong, P.; Taillefer, D.; Laking, J.; Pineault, J.; Chader, G.; Tennisswood, M.  
Eur. J. Biochem. 221, 917-925, 1994  
A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with  
A:Reference number: S43646; MUID:94237156  
A:Accession: S43646  
A:Molecule type: DNA  
A:Residues: 1-449 <WON>  
A:Cross-references: GB:M64722; NID:9339972; PIDN:AAB06508.1; PID:9339973  
R:Kirschbaum, L.; Sharpe, J.A.; Murphy, B.; d'Aplice, A.J.F.; Classon, B.; Hudson, P.;  
EMBO J. 8, 711-718, 1989  
A:Title: Molecular cloning and characterization of the novel, human complement-associ  
A:Reference number: S04662; MUID:89251601  
A:Accession: S04662  
A:Molecule type: mRNA  
A:Residues: 1-449 <KIR>  
A:Cross-references: EMBL:X14723; NID:930250; PIDN:CAA32847.1; PID:930251  
A:Note: parts of this sequence, including the amino end of the mature protein, were c  
R:Jenne, D.E.; Tschopp, J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989  
A:Title: Molecular structure and functional characterization of a human complement cy  
lis fluid.  
A:Reference number: A41386; MUID:89386692  
A:Accession: A41386  
A:Molecule type: mRNA  
A:Residues: 2-449 <JEN>  
A:Cross-references: GB:M25915; NID:9180619; PIDN:AAA51765.1; PID:9180620  
R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J.  
Biochemistry 29, 5380-5389, 1990  
A:Title: Apolipoprotein J: structure and tissue distribution.  
A:Reference number: A35833; MUID:90344779  
A:Accession: A35833  
A:Molecule type: mRNA  
A:Residues: 34-449 <DES>  
A:Cross-references: GB:J02908; NID:9178854; PIDN:AAA51765.1; PID:9178855  
R:Gloss, J.; Matsubara, E.; Koudinov, A.; Choi-Miyura, N.H.; Tomita, M.; Wisniewski, T  
Biochem. J. 293, 27-30, 1993  
A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's  
A:Reference number: S34056; MUID:93319521  
A:Accession: S34056  
A:Molecule type: protein  
A:Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GH>  
R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse  
Arterioscler. Thromb. 11, 645-652, 1991

A:Title: Characterization of a human high density lipoprotein-associated protein, NA1/NA  
A:Reference number: A53177; MUID:91230083  
A:Accession: A53177  
A:Molecule type: protein  
A:Residues: 229-242:303-304, 'M', 306-312, 'X', 314-317,397-403 <IAM>  
R:de Silva, H.V.; Stuart, M.D.; Park, Y.B.; Mo, S.O.T.; Gill, C.M.; Wetterau, J.R.; Busc  
J. Biol. Chem. 265, 14292-14297, 1990  
A:Title: Purification and characterization of apolipoprotein J.  
A:Reference number: A37816; MUID:90354412  
A:Accession: A37816  
A:Molecule type: protein  
A:Residues: 23-46, 'H', 48-51, 'Q', <DE3>  
A>Note: amino end of the alpha chain  
A:Accession: B37816  
A:Molecule type: protein  
A:Residues: 220-257 <DE2>  
A>Note: amino end of the beta chain  
R:Choi, N.H.; Mazda, T.; Tomita, M.  
Mol. Immunol. 26, 835-840, 1989  
A:Title: A serum protein SP40,40 modulates the formation of membrane attack complex of C  
A:Reference number: PLO135; MUID:90097955  
A:Accession: PLO136  
A:Molecule type: protein  
A:Residues: 23-37 <CHO>  
A>Note: this fragment was isolated from the membrane attack complex SC5b-9  
A:Accession: PLO135  
A:Molecule type: protein  
A:Residues: 228-242 <CH2>  
A>Note: this fragment was isolated from the membrane attack complex SC5b-9  
R:Hochstrasser, A.C.; James, R.W.; Martin, B.M.; Hochstrasser, D.; Pomet  
Appl. Theor. Electrophor. 1, 73-76, 1988  
A:Title: HDL particle associated proteins in plasma and cerebrospinal fluid: identificat  
A:Reference number: S07433; MUID:91265608  
A:Accession: S09339  
A:Molecule type: protein  
A:Residues: 229-240 <HOC>  
A:Accession: S07433  
A:Molecule type: protein  
A:Residues: 24-27, 'S', 29-33 <HOC>  
R:Kirsbaum, L.; Bozas, S.E.; Walker, I.D.  
FEBS Lett. 297, 70-76, 1992  
A:Title: SP-40,40, a protein involved in the control of the complement pathway, possesses  
A:Reference number: A56293; MUID:92201397  
R:Kunitake, S.T.; Carilli, C.T.; Lau, K.; Procter, A.A.; Naya-Vigne, J.; Kane, J.P.  
Biochemistry 33, 1988-1993, 1994  
A:Title: Identification of proteins associated with apolipoprotein A-I-containing lipop  
A:Reference number: A54223; MUID:94162201  
A:Accession: E54223  
A:Molecule type: protein  
A:Residues: 228-246 <KUN>  
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma  
A:Accession: F54223  
A:Molecule type: protein  
A:Residues: 23-34, 'X', 36-37 <KU2>  
A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma  
R:Danik, M.; Chabot, J.G.; Mercier, C.; Benbidid, A.L.; Chauvin, C.; Quirion, R.; Sub, M.  
Proc. Natl. Acad. Sci. U.S.A. 88, 8577-8581, 1991  
A:Title: Human gliomas and epileptic foci express high levels of a mRNA related to rat t  
A:Reference number: I59206; MUID:92020896  
A:Accession: I59206  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 61-449 <RES>  
A:Cross-references: GB:M74816; NID:g338056; PIDN:AAA60321.1; PID:g338057  
R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtelotte, W.W.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989  
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.  
A:Reference number: I48174; MUID:89386721  
A:Accession: I63132  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 103-168 <RE2>

A:Cross-references: GB:M26639; NID:g338070; PIDN:AAA36609.1; PID:g553644  
C:Comment: This protein has been implicated in complement cascade inhibition, membran  
11 as in normal brain. In tissues affected by neurodegenerative disease processes, an  
C:Comment: This protein may assist in preventing the formation of Alzheimer's disease  
C:Genetics:  
A:Gene: GDB:CLU; CLU  
A:Cross-references: GDB:125226; OMIM:185430  
A:Map position: 8p21-8p21  
A>Note: appears to be a single-copy gene; alternative exon usage in 5'-untranslated r  
C:Superfamily: clusterin  
C:Keywords: apoptosis; complement inhibitor; extracellular protein; glycoprotein; HDL  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-227/Domain: clusterin beta chain #status experimental <BCH>  
F:23-227,228-449/Product: clusterin #status experimental <MAT>  
F:228-449/Domain: clusterin alpha chain #status experimental <ACH>  
F:86,103,145,291,354,374/Binding site: carbohydrate (asn) (covalent) #status experime  
F:102-313,121-295,129-285/Disulfide bonds: #status experimental  
F:113-305,116-302/Disulfide bonds: (or 113-302, 116-305) #status experimental  
F:317/Binding site: carbohydrate (asn) (covalent) #status absent

Query Match 17.2%; Score 434; DB 1; Length 449;  
Best Local Similarity 25.8%; Pred. No. 4.3e-20;  
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

```
QY 12 MKPPLVFIYVCLMLKDSHCAPTWK-----DKTAISNLSKFSFVSGIDADEYKALT 65
DB 1 MKKTLILFVGLL-----TWESGYLDQYVSDNELDMSQSKYNNKELONVNV 51
QY 66 GIKOKIMMERKEKHTLMSTLKKREKQDAKLNEVOHEEERLCRESIASWG 125
DB 52 GVKQIKLTIEKTNEBRKTLNLNEAKKKEDALNETRESFTKLKELPGVCEETMALME 111
QY 126 ECRSLENNCKRIT-TCQPSMSYKNIIEFFKITYFLPPHEDNEKDLPISEKLEK 184
DB 112 ECKPLKQTKMFKYARVCRSSGLVGRQLEELFNQSSPFYMMNGDR-----IDSLLEN 165
QY 185 DAOLTQMEDV----FSQTLTVNLSLFNRSFNVFMQMOEFOQTQSPHSIS-DTDLTEPYF 239
DB 166 DRQOTHMVDVQMHDSKSSIIIDELFQDRF-----FTREPDYT--HLPSPSLHRRPHF 218
QY 240 FPAFSKEMPTADLE-QCWDIPNFQFLCNPSVSIYESVETTKMLKAIEDLPKODKAP 298
DB 219 F--FPKSRIVSLMFPFSYEPLNFHAMQPLEIMEA-----QOAMDIFHSPAF 267
QY 299 DHGGLISMLRGDGLGELDONLSRCFKEHEKCOQANLSBDC-----PDVPAHLTEL 354
DB 268 QHPTEPIREGDDBRYVCREIRHNSITGCLRNKDKDCREILSVDCSTNNSQAKLRREL 327
QY 355 DEAIRLVNVSQOYGOALQMTKRLHLEDYAVLVKMRGQFGVNSLANQAPETEIFNSIQ 414
DB 328 DESIQVABRLRKYNELKSYQWKMILNSSLDELQNGFNVSFLANLQGEDEQYILRVY 387
QY 415 VVPRTHGEGNISKQETMTDLSIPSSNFTKIDLEBSAESSNFTGYVAKALQHF-KEH 473
DB 388 TVAS-HRSDSDPVSGVTEVVVKLEFSDSPITVTVPEVVSRRKNPKMETVAEKALQDEYRKH 446
```

RESULT 2  
clusterin - quail  
C:Species: Coturnix coturnix (quail)  
C:Date: 13-Sep-1996 #sequence\_revision 13-Sep-1996 #text\_change 13-Aug-1999  
C:Accession: I50131  
R:Michel, D.; Chatalein, G.; Herault, Y.; Brun, G.  
Eur. J. Biochem. 229, 215-223, 1995  
A:Title: The expression of the avian clusterin gene can be driven by two alternative  
A:Reference number: I50131; MUID:95262670  
A:Accession: I50131  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-451 <MIC>  
A:Cross-references: EMBL:X80760; NID:g520629; PIDN:CAA56733.1; PID:g520630

C:Genetics:  
A:Gene: T64  
A:introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2  
C:Superfamily: clusterin

Query Match 16.5%; Score 418; DB 2; Length 451;

Best Local Similarity 25.7%; Pred. No. 4.5e-19;  
Matches 118; Conservative 94; Mismatches 179; Indels 68; Gaps 15;

```

QY 45 LKSFSEVGEIDADEYKKAALTGKIMMERKEKHTNIMSLTKCREKQKALALNE 104
DB 27 LKQLSAAGSKYIDAENVAINGVKMKTLMDKSKHEQAMLHLETKKKEAVKLALE 86
QY 105 VOHLEEEERLCRESLADSWGECRSCLENNCMRIYT-TCOPSSSVYKNIERFFKRIYOF 163
DB 87 KEKQLAEKQEVCEHTMLSLMEBECKPCLKHCAMVYSKMGHSSGLVGRQLEELLNSSPF 146
QY 164 LPPFHEDNEKDLPISEKLEK-----DAQLTQMEDVFSQLTVDVNSLFNRSFNVPRQMOQ 218
DB 147 SIWNGERIDLDLRQROGERREFDELEERFGLMEDGVEDIFQDSTQLYGPAFFPFR---- 202
QY 219 EFDQTFQSHFISDTDLTEPY--FFPAFSKPEPMTKADLEQCWDIPN-----FFQLRC 267
DB 203 -----TPPGGFRFAF-VPPVQKVAH-----VPRRLSLREHLPFRQHPM 240
QY 268 NFSVSIYESVETITKMLKAIEDLPKQDKAPDH--GGLISKMLP-GQDRGLGCLDQNLTS 324
DB 241 HGHRHLPFLPFLFEMTQMLDG-----GHGMEHPLGFPATESNFTDRMYCEIRNNSA 294
QY 325 RCFKFEKQCKQCAHLSEDC---PVPALHLEDAIRLVNSNOQYGILOMTRKHLE 380
DB 295 GCLRMDECEKCEKREILAVDCSQTPVQSOLRQEFEDALRIAEFRTRYDILLSAFAEML 354
QY 381 DTAFLVKKMGSGWVSELANQAP-----ETELIFNSIQVVRHIEGNISKODETMMD 434
DB 355 NTSLLDQDLARQFGWVSRIGNLTGQNDGLQVTTVFESK--TNLEDP--SAPADTQVLT 408
QY 435 LSTLPSSNFTLKIPLSESAESSNFIQYVAKALQHFKEH 473
DB 409 VQLFDESEPLSLTYPGDISMDPRFMEIVAEQALQHYKON 447

```

# RESULT 3

clusterin precursor - dog  
N:Alternate names: 80k secreted glycoprotein, renal; complement cytotoxic inhibitor

C:Species: Canis lupus familiaris (dog)  
C>Date: 20-Mar-1992 #sequence-revision 20-Mar-1992 #text-change 13-Aug-1999

C:Accession: A40018  
R:Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pillarsky, C.; Appel, D.; Ha

J. Biol. Chem. 266, 9924-9931, 1991  
A:Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in

A:Reference number: A40018; MUID:91236776

A:Accession: A40018  
A:Molecule type: mRNA

A:Residues: 1-445 <HAR>  
A:Cross-references: GB:M55251; GB:M38757; NID:9163953; PIDN:AAA30846.1; PID:9163954

A:Note: Parts of this sequence, including the amino ends of the alpha and beta chains, w

C:Superfamily: clusterin  
C:Keywords: apoptosis; disulfide bond; glycoprotein

F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-226/Domain: clusterin beta chain #status experimental <BCH>

F:23-226/227-445/Product: clusterin #status experimental <MAT>  
F:227-445/Domain: clusterin alpha chain #status experimental <CH>

F:86,103,145,277,287,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 16.2%; Score 410.5; DB 2; Length 445;  
Best Local Similarity 23.9%; Pred. No. 1.3e-18;  
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 12 MKPPLVFIYCLIMLKDSHCAPTWKDKTAISENLSFSEVGEIDAD-----EEVKKALT 65

```

DB 1 MKKTLILLVGLL-----TWNGRVLGQAVSDVTELOEMSTEGSKYINKEIKNAK 51
QY 66 GIKOKIMMERKEKHTNIMSLTKCREKQKALALNEVOHLEEEERLCRESLADSWG 125
DB 52 GKQKITLIEQNEERKSLISLNEEAKKKEDALNDTKSEFKLASQGVCDTMAALME 111
QY 126 ECRSCLENNCMRIYT-TCOPSSSVYKNIERFFKRIYOFLPPFHEDNEKDLPISEKLEK 184
DB 112 ECKPCLQTKQCMFYARVCSGSGLVGHQLEEFNLNOSPFPYFMWNGDR-----IDSLLEN 165
QY 185 DAQLTQMEDVFSQLTVDVNSLFNRSFNVPRQMOQEDTFQSHFIS---DTDLTEPYFE 240
DB 166 DRQQTALDY-----MODSFRASSI-----MDELFORFFRFPDPTVHYSP--F 209
QY 241 PAFSKPEPMTKADLEQCWDI-----PNEFOLCFNFSVSIYESVETITKMLKAIE--- 289
DB 210 SLFORRPFPPKFRARIANIIPPRFQPLNFHDFQFFPMIQA-QQAMDVNLHRIPIYHF 268
QY 290 --DLPKQDKAPDHGGLISKMLPQDRGLGCLDQNLISRCFKFEKQCKQCAHLSEDC--- 344
DB 269 PLIEPFEED-----NRTVCKEIRHNSTGCLKMKDKQCEKCOELLISVDCSSN 312
QY 345 -PVPALHLEDAIRLVNSNOQYGILOMTRKHLEDAIRLVNVEKKRGQFGVSELANQA 403
DB 313 NPAQVQLRQELNSQLIAEKFTKLYDELQSYQEKMFNTSSLKQLEQFSWVSQLANLT 372
QY 404 PETELIFNSIQV-PRHIEGNISKODETMMDTSLPSSNFTLKIPLSESAESSNFIQYV 462
DB 373 QSEDFPYLTQVTTVGSOTSDSNVPVGFYKYV--VKLDSQPTIVMIPAVNSRNNPFMEIV 430
QY 463 VAKALQHFKE 472
DB 431 AEKALQGYRQ 440

```

RESULT 4  
S07714  
T64 protein precursor - Japanese quail

C:Species: Coturnix coturnix japonica (Japanese quail)  
C>Date: 29-Jan-1993 #sequence-revision 29-Jan-1993 #text-change 13-Aug-1999

C:Accession: S07714  
R:Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.

A:Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b

A:Reference number: S07714; MUID:89239492

A:Accession: S07714  
A:Molecule type: mRNA

A:Residues: 1-449 <MIC>  
A:Cross-references: EMBL:X15825; NID:962594; PIDN:CAA3823.1; PID:962595

C:Superfamily: clusterin  
F:1-18/Domain: signal sequence #status predicted <SIG>

F:19-449/Product: T64 protein #status predicted <MAT>

Query Match 16.1%; Score 408; DB 2; Length 449;  
Best Local Similarity 25.7%; Pred. No. 1.9e-18;  
Matches 117; Conservative 92; Mismatches 179; Indels 68; Gaps 15;

QY 45 LKSFSEVGEIDADEYKKAALTGKIMMERKEKHTNIMSLTKCREKQKALALNE 104

DB 27 LKQLSAAGSKYIDAENVAINGVKMKTLMDKSKHEQAMLHLETKKKEAVKLALE 86

QY 105 VOHLEEEERLCRESLADSWGECRSCLENNCMRIYT-TCOPSSSVYKNIERFFKRIYOF 163

DB 87 KEKQLAEKQEVCEHTMLSLMEBECKPCLKHCAMVYSKMGHSSGLVGRQLEELLNSSPF 146

QY 164 LPPFHEDNEKDLPISEKLEK-----DAQLTQMEDVFSQLTVDVNSLFNRSFNVPRQMOQ 218

DB 147 SIWNGERIDLDLRQROGERREFDELEERFGLMEDGVEDIFQDSTQLYGPAFFPFR---- 202

QY 219 EFDQTFQSHFISDTDLTEPY--FFPAFSKPEPMTKADLEQCWDIPN-----FFQLRC 267





RESULT 7  
A35744  
Clusterin precursor - bovine  
N:Alternate names: complement cytolysis inhibitor; glycoprotein III  
M:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 13-Aug-1999  
A:Accession: A35744  
R:Palmer, D.J.; Christie, D.L.  
J. Biol. Chem. 265, 6617-6623, 1990  
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin cells  
A:Reference number: A35744; MUID:90216681  
A:Accession: A35744  
A:Molecule type: mRNA  
A:Residues: 1-439 <PAL>  
A:Cross-references: GB:J05391; NID:9163114; PIDN:AAA30554.1; PID:9163115  
A:Note: Parts of this sequence, including the amino ends of the alpha and beta chains, were derived from complementary DNA clones.  
C:Superfamily: clusterin  
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-221/Domain: clusterin beta chain #status experimental <BCH>  
F:20-221,222-439/Product: clusterin #status experimental <MAT>  
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>  
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.8%; Score 400.5; DB 2; Length 439;  
Best Local Similarity 24.5%; Pred. No. 5.5e-18;  
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;  
34 TWKDKTAISE-NLSESEVEIDADEEVKALTGKIMKIMREKREHTNLMTLTKCR 92  
13 SWMSGNAISKEQEMSTESKRYVNEIKALKKEVQIKQIOTNEERKLLSSLEAK 72  
93 EERQALKLINEVOHLEEEERLCRESLADSGECSLENNCMRIYT-TCQPSWSSVKN 151  
73 KKREDAINDRDESENKLASQGVNETMTALMECKPCLQTCMKFYARVCRSGSLVGH 132  
152 KIRFRKRIQYLFPPFHEDNEKDLPISEKLEDAQLOQMEDVFSQTVNLSLFRSN 211  
133 QLEEFINQSSPFYFWINGDR-----IDSLMENDREQSHMDV-----MEDSFRRASS 179  
212 VFRQMOQEFQOTQSHFI-----SDTDLTEPY-----FF--PAFSKEPMTKADLEQC 256  
180 I-----MDLFDQDRFLRRPDQYQYSPSSFRGSLFFNPKSRARAVMFPPLLEP 232  
257 WDIPNFFQLFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPRHGGLISMLPGQDRGLC 316  
233 ----NFHDVQPF-----YDMIHQAQAMDAHLQ-----RTPYHFPFM-EFTENNRTVC 277  
317 GELDONLSRQCFKHEKQCOQAHLSEDC-----PDVPAHTELDLAIPLVAVNSQOQIL 372  
278 KETRNHSTGCLRKQDCEKQETLEVDCSANPQIILLRQQLNLSQDLAKFSRLDQL 337  
373 QMTRKHLDEATYLVKMRGQFGWVSELANQAPETE-----IIFNSIQVVPRIHEGNISK 436  
338 OSVQKMLNLSALLKQNEQFTWSQLANTQSDQHYLQVFNNSNPSPLSG----- 393  
427 QDEMTATDLSILPSSNTTLKIPLESAESSNFICYVAAKLQHRK 471  
394 ----LTKVIYKLENSPPIITVTVQAEVSSPNFMENVAEKALQOYR 433

RESULT 8  
A42108  
Clusterin precursor - pig  
N:Alternate names: complement cytolysis inhibitor; CP40 protein  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
A:Accession: A42108; J05535; PC4475  
R:Demmer, V.; Hoyle, M.; Baglioni, C.; Mills, A.J.  
J. Biol. Chem. 267, 5237-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic  
A:Reference number: A42108; MUID:92184774  
A:Accession: A42108  
A:Molecule type: mRNA  
A:Residues: 1-446 <DIE>  
A:Cross-references: GB:M84639; NID:9164408; PIDN:AAA31013.1; PID:9164409  
A:Experimental source: aortic smooth muscle cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)  
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K  
Biochem. Biophys. Res. Commun. 234, 712-718, 1997  
A:Title: The glycoproteins that occur in the colligoids of senescent porcine pituitary  
A:Reference number: J05535; MUID:97318844  
A:Accession: J05535  
A:Molecule type: DNA  
A:Residues: 1-446 <OGA>  
A:Accession: PC4475  
A:Molecule type: protein  
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG>  
A:Experimental source: pituitary gland  
C:Superfamily: clusterin  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-227/Domain: clusterin beta chain #status predicted <BCH>  
F:23-227,228-446/Product: clusterin #status predicted <MAT>  
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

Query Match 15.8%; Score 400.5; DB 2; Length 446;  
Best Local Similarity 24.4%; Pred. No. 5.6e-18;  
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;  
16 LWFVYCLMLKDSHCAPTWK-----DKTAISENLKSEVEGIDADEEVKALTGIR 68  
4 LLLVLVGLL-----TWENGPVYLGDKAISDKELQEMSTESKRYVNEIKALKLEVK 54  
69 QMKIMREKREHTNLMTLTKCREKQALKLINEVOHLEEEERLCRESLADSGEGR 128  
55 QIKTLQSGNEERKSLSSLEAKKKEDALNDTRTEKRLGSGQLCNETMALMECK 114  
129 SCLENNCMRIYT-TCQPSWSSVKNKIERFRKRIQYLFPPFHEDNEKDLPISEKLEKDAQ 187  
115 PCLKQTCMKFYARVCRSGSLVGHQLEEFINQSSPFYFWINGDR-----IDSLMENDQ 168  
188 LQMEDVFSQTVNLSLFRSNFNRQMOQEFQOTQSHFIS-----DITDLTEPY----- 238  
169 QSHVMDI-----MEDSFRRASSNI-----MDLFDQDRFLRRPDQYQYSPSSFRGSLFFNPKSRARAVMFPPLLEP 214  
239 ----FFPAFSKEPMTKADLEQCDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKQD 295  
215 GSLFFNPKSRFARNIMPFLETDL-NYHDMFQPF-----EDMTHQAQAMDAHLRIPIYH- 268  
296 KAPDHGGLISKMLP--GDRGLCGELDONLSRQCFKHEKQCOQAHLSEDC-----PDVPA 349  
269 -FPEAG-----VPENSNRAVCKEIRHNSSTGCLRKQDCEKQRELSSQCSANSSQMO 321  
350 LHTLEDLAIPLVAVNSQOQIILQMTKHLDEATYLVKMRGQFGWVSELANQAPETEIT 409  
322 LRQELTSLQMAKFSKLYDQILQSQKMLNLSLKLQLENEFVNSQLANTQDNDY 381  
410 FNSIQVVPRIHEGNISKQDETMTDLSILPSSNTTLKIPLESAESSNFICYVAAKLQ 469  
382 YLDVTVV-NSHGSDDPSVPSGLTVVVKLFDSYITLILQFVYS--DPKFMETVAEBALQ 438  
470 FKE 472  
439 YRQ 441

RESULT 9  
A27205  
Clusterin precursor - rat  
N:Alternate names: SGP-2; sulfated glycoprotein 2; TRPM-2  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 13-Aug-1999

C:Accession: A45890; S18491; A45415; B31575; A31575; A27205  
R:Collard, M.W.; Griswold, M.D.  
Biochemistry 26, 3297-3303, 1987  
A:Title: Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat ;  
A:Reference number: A45890; MUID:88000523  
A:Accession: A45890  
A:Molecule type: mRNA  
A:Residues: 1-447 <C02>  
A:Cross-references: GB:M16975; NID:g204472; PIDN:AAA41273.1; PID:g204473  
R:Bettermann, S.; Hlupka, R.A.; Gilna, P.; Liao, S.  
Biochem. J. 257, 293-296, 1989  
A:Title: Identification of an androgen-repressed mRNA in rat ventral prostate as coding  
A:Reference number: S18491; MUID:89149740  
A:Accession: S18491  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-186, 'D', 188-447 <BET>  
A:Cross-references: EMBL:X13231; NID:957240; PIDN:CAA31610.1; PID:957241  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988  
J: Wong, P.; Pineault, J.; Lelands, J.; Taillefer, D.; Leger, J.; Wang, C.; Tenniswood, M.  
J. Biol. Chem. 268, 5021-5031, 1993  
A:Title: Genomic organization and expression of the rat PRP-M2 (clusterin) gene, a gene  
A:Reference number: A45415; MUID:93186813  
A:Accession: A45415  
A:Status: preliminary  
A:Molecule type: DNA; mRNA  
A:Residues: 1-447 <MON>  
A:Experimental source: blood  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBI:126803, NCBI:126805)  
R:Cheng, C.Y.; Chen, C.L.C.; Feng, Z.M.; Marshall, A.; Bardin, C.W.  
Biochem. Biophys. Res. Commun. 155, 398-404, 1988  
A:Title: Rat clusterin isolated from primary Sertoli cell-enriched culture medium is sul  
A:Reference number: A90146; MUID:88326333  
A:Accession: B31575  
A:Molecule type: protein  
A:Residues: 22-51 <CHD>  
A:Note: amino end of the beta chain  
A:Accession: A31575  
A:Molecule type: protein  
A:Residues: 227-256 <CH2>  
A:Note: amino end of the alpha chain  
C:Superfamily: clusterin  
C:Keywords: Apoptosis; disulfide bond; glycoprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-226/Domain: clusterin beta chain #status experimental <BCH>  
F:227-447/Domain: clusterin alpha chain #status experimental <MA>  
F:102-144,290,327,353/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match	15.08	Score 380	Db 2	Length 447
Best Local Similarity	23.55	Pred. No. 1,1e-16		
Matches 113	Conservative 103	Mismatches 197	Indels 68	Gaps

  

QY	16	L L V E T V C L L M L K D S H C A P T M W K D K T A I S E N L K S F S E V E I D - - - - - A D E E Y K A L T I Q I K Q	69
Db	4	L L L C V A L L L - - - - - T W D N G A V L G E P F S D N E L G E L S T Q S R Y Y N K E I Q N N A V G V K H	54
QY	70	M K I M M E R K E K E N T M L M S T I K K C R E E K Q E A L K L N E V O D H L E E E R L C R E S T I A D S W G C R S	123
Db	55	I K T I E K T M E R K S L N S L E A K K K K G A L D D T R D S M K K A P E V C N E T M M L M E C K P	114
QY	130	C L E N N C K A I T Y - T C O P S W S S Y K N I E R F F R K I Y O L F P P H E D N K D L P I S E K L I E K A O L	188
Db	115	C L K H T C K K F A R A V C R S S G L G R Q L E E F L N S S P F Y T M M N G D R - - - - - I D S L L E S R Q Q	168
QY	189	T Q - - - - - M E D V S Q L T V D V N S L F P N S E V F E R M O D E P Q T O S H E I S D T L D T E Y F F P - -	241
Db	169	S Q V L D A M O D S T F R A S G I I L F O D R F - F T H E P O D I - - - - - H H F - - - - - S P M G F P H K R	214
QY	242	- - - - - A F S K E P M T K A D L E G C W D I P - N F O L F C N F S V I S E S V S E T I T K M L K A I E D L P R O D K A	297

Db	215	PHFLYKSRILVRSLMPLSHVBERLSEFNHMGQFFPMIHOQAQAMVOYLHSPALQEPVDEL	274
QY	298	PDHGGLSKMLPGQ-DRLGCELDONISRCFKFHEKQCOAKHLEDSC----	PDVPLATT 352
Db	275	KE-----GDDPFWCKEIRHNSTGCLKMKGOCEKCECILSYVDCSTNNPQAANLRQ	324
QY	353	ELDEAIRLVNVSNOOYGILOMTFRKHLDPYAVLEKKMRGOGWVSELAN--QAPEFTIF	410
Db	325	ELNDSLOVAERLTOYINELLHSLQSKMLNVSLLLEQDLQDTWVSQALANTQGGDOYLRY	384
QY	411	NSIOVPRIRHEGNISKQDETMATDSLIPSSNFTLPIPEESASSNFICYVAKALOHF	470
Db	385	STVTT---HSSDESEVSRVTEVVVKLFEDSPITVIVLPEEVSXKONPKFMDTVAEKALQEX	440
QY	471	K 471	
Db	441	R 441	

RESULT	10
148174	
sulfated glycoprotein 2 - golden hamster (fragment)	
C:Species: Mesocricetus auratus (golden hamster)	
C:Date: 27-Feb-1997 #sequence_revision 27-Feb-1997 #text_change 13-Aug-1999	
C:Accession: 148174	
P:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtelotte, W.W.	
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989	
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease	
A:Reference number: 148174; MUID:89386721	
A:Accession: 148174	
A:Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: mRNA	
A:Residues: 1-191 <RES>	
A:Cross-references: GB:M26640; NID:g191450; PIDN:AAA37102.1; PID:g191451	
A:Superfamily: Clusterin	

[illegible]

RESULT	11
Interactin - slime mold (Dictyostelium discoideum)	
C:Species: Dictyostelium discoideum	
C:Date: 20-Sep-1999 #sequence_revision 20-Sep-1999 #text_change 11-May-2000	
C:Accession: T14867	
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.	
J. Cell Biol. 142, 735-750, 1998	
A:Title: Interactin, an actin-binding protein of the alpha-actinin superfamily in Dicty.	
A:Reference number: Z18248; MUID:98365468	
A:Accession: T14867	
A>Status: preliminary; translated from GB/EMBL/DBJ	
A:Molecule type: DNA	
A:Residues: 1-1738 <RIV>	
A:Cross-references: EMBL:AF057019; NID:93549260; PID:93549261; PIDN:AAC34582.1	
C:Genetics:	
A:Gene: abpD	
A:Introns: 173/2; 1680/1	
Query Match	6.5%; Score 164; DB 2; Length 1738;
Best Local Similarity	21.1%; Pred. No. 0.025;





Search completed: July 2, 2002, 11:52:02  
Job time: 225 sec

---

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:58:14 ; Search time 17.57 Seconds  
(without alignments)  
1051.180 Million cell updates/sec

Title: US-09-722-544A-4MOD

Perfect score: 2529

Sequence: 1 MRTWYSNSGNMKPELLVFI.....FIGYVAKALQHFKEFTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	434	17.2	449	1	CLUS_HUMAN
2	418	16.5	451	1	CLUS_COTJA
3	413.5	16.4	447	1	CLUS_RABIT
4	410.5	16.2	445	1	CLUS_CANFA
5	403	15.9	448	1	CLUS_MOUSE
6	400.5	15.8	439	1	CLUS_BOVIN
7	400.5	15.8	446	1	CLUS_PIG
8	380	15.0	447	1	CLUS_RAT
9	158	6.6	191	1	CLUS_MESAU
10	151	6.0	1526	1	MY52_SCHPO
11	145.5	5.8	594	1	CIK1_YEAST
12	145.5	5.8	1938	1	MYHD_HUMAN
13	141.5	5.6	1005	1	RA50_METJA
14	141.5	5.6	3433	1	UTRO_HUMAN
15	139.5	5.5	2653	1	GENE_HUMAN
16	138	5.5	5430	1	ACR7_HUMAN
17	137	5.4	1940	1	MYH3_HUMAN
18	137	5.4	2230	1	GOG4_HUMAN
19	135.5	5.4	1679	1	YIO9_YEAST
20	135	5.3	1060	1	EG51_XENLA
21	133.5	5.3	1957	1	YD86_SCHPO
22	132.5	5.2	852	1	RA50_THEMA
23	131.5	5.2	1978	1	MYHB_CHICK
24	130.5	5.2	1940	1	MYH3_RAT
25	130	5.1	793	1	KATP_ARATH
26	129.5	5.1	1427	1	REST_HUMAN
27	129.5	5.1	2077	1	TEGU_HSV6U
28	129	5.1	539	1	MY53_HYDAT
29	128	5.1	1939	1	MYH1_HUMAN
30	128	5.1	1744	1	TANA_XENLA
31	128	5.1	1853	1	MY5A_MOUSE
32	128	5.1	2704	1	BPAL_HUMAN
33	127.5	5.0	722	1	MFPL_TOBAC

34	127	5.0	1189	1	SCII_CHICK	Q90988 gallus gall
35	126.5	5.0	971	1	Y228_BORBU	O51246 borrelia bu
36	126.5	5.0	1875	1	MLPI_YEAST	Q02455 saccharomyc
37	126	5.0	1941	1	MYH2_HUMAN	Q9ukx2 homo sapien
38	126	5.0	1966	1	MY5B_CAEEL	P02566 caenorhabdi
39	125.5	5.0	978	1	RA5Q_AQAE	O67124 aquifex aeo
40	125.5	5.0	1067	1	EG52_XENLA	Q91783 xenopus lae
41	125.5	5.0	1969	1	MYSA_CAEEL	P12844 caenorhabdi
42	125.5	5.0	3210	1	CENP_HUMAN	P49454 homo sapien
43	125	4.9	1790	1	USO1_YEAST	P25386 saccharomyc
44	125	4.9	1960	1	MYH9_HUMAN	P35579 homo sapien
45	125	4.9	2418	1	SPCA_HUMAN	P02549 homo sapien

## ALIGNMENTS

RESULT	1
CLUS_HUMAN	
ID	CLUS_HUMAN STANDARD; PRT; 449 AA.
AC	P10309; P11380; P11381;
DT	01-JUL-1989 (Rel. 11, Created)
DT	01-JUL-1989 (Rel. 11, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Clusterin precursor (Complement-associated protein SP-40,40)
DE	(Complement cytolysis inhibitor) (CLI) (NAL and NA2) (Apolipoprotein J) (Apo-J) (TRPM-2).
GN	CLU.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Liver;
RX	MEDLINE=89251601; PubMed=2721499;
RA	Kirsbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,
RA	Hudson P., Walker I.D.;
RT	"Molecular cloning and characterization of the novel, human
RT	complement-associated protein, SP-40,40: a link between the
RT	complement and reproductive systems.";
RL	EMBO J. 8:711-718(1989).
RN	[2]
RP	SEQUENCE OF 2-449 FROM N.A.
RX	MEDLINE=89386692; PubMed=2780565;
RA	Jenne D.E., Tschopp J.;
RT	"Molecular structure and functional characterization of a human
RT	complement cytolysis inhibitor found in blood and seminal plasma:
RT	identity to sulfated glycoprotein 2, a constituent of rat testis
RT	fluid.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=90344779; PubMed=1974459;
RA	de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;
RT	"Apolipoprotein J: structure and tissue distribution.";
RL	Biochemistry 29:5380-5389(1990).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94237156; PubMed=8181474;
RA	Wong P., Taillefer D., Lakin J., Pineault J., Chader G.,
RT	Tennisswood M.;
RT	"Molecular characterization of human TRPM-2/clusterin, a gene
RT	associated with sperm maturation, apoptosis and neurodegeneration.";
RL	Eur. J. Biochem. 221:917-925(1994).
RN	[5]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=93186813; PubMed=7680346;
RA	Pineault J.M., Tennisswood M.;
RT	"Genomic organization and expression of the rat TRPM-2 (clusterin)
RT	gene, a gene implicated in apoptosis.";
RL	J. Biol. Chem. 268:5021-5031(1993).
RN	[6]

RP SEQUENCE FROM N.A.  
RA Blechschmidt K., Rosenthal A.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 61-449 FROM N.A.  
RX MEDLINE=92020896; PubMed=1924317;  
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,  
RL Quirion R., Suh M.;  
RT "Human gliomas and epileptic foci express high levels of a mRNA  
related to rat testicular sulfated glycoprotein 2, a purported marker  
of cell death.";  
RT Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).  
RN [8]  
RP SEQUENCE OF 140-449 FROM N.A.  
RC TISSUB=Liver;  
RA Glew M.D., Kirszbaum L., Bozas S.E., Walker I.D.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP SEQUENCE OF 23-33 AND 228-240.  
RX MEDLINE=91265608; PubMed=3154963;  
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,  
RL Hochstrasser D.F., Pometta D., Merril C.R.;  
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:  
identification and partial sequencing.";  
RL Appl. Theor. Electrophor. 1:73-76(1988).  
RN [10]  
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.  
RX MEDLINE=91230083; PubMed=1903064;  
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,  
RL Hochstrasser D.F.;  
RT "Characterization of a human high density lipoprotein-associated  
protein, NAI/NA2. Identity with SP-40,40, an inhibitor of complement-  
mediated cytolysis.";  
RL Arterioscler. Thromb. 11:645-652(1991).  
RN [11]  
RP SEQUENCE OF 23-52 AND 228-257.  
RX MEDLINE=90354412; PubMed=2387851;  
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,  
RL Wetterau J.R., Busch S.J., Harmony J.A.K.;  
RT "Purification and characterization of apolipoprotein J.";  
RL J. Biol. Chem. 265:14292-14297(1990).  
RN [12]  
RP SEQUENCE OF 23-37 AND 228-242.  
RX MEDLINE=90097955; PubMed=2601725;  
RA Choi N.H., Mazda T., Tomita M.;  
RT "A serum protein SP40,40 modulates the formation of membrane attack  
complex of complement on erythrocytes.";  
RL Mol. Immunol. 26:835-840(1989).  
RN [13]  
RP SEQUENCE OF 23-41 AND 228-246.  
RX MEDLINE=93319521; PubMed=8328966;  
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,  
RL Wisniewski T., Frangione B.;  
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is  
complexed to SP-40,40 (apolipoprotein J), an inhibitor of the  
complement membrane-attack complex.";  
RL Biochem. J. 293:27-30(1993).  
RN [14]  
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=92201397; PubMed=1551440;  
RA Kirszbaum L., Bozas S.E., Walker I.D.;  
RL "SP-40,40, a protein involved in the control of the complement  
pathway, possesses a unique array of disulphide bridges.";  
RL FEBS Lett. 297:70-76(1992).  
RN [15]  
RP DISULFIDE BONDS.  
RX MEDLINE=93147076; PubMed=1491011;  
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;  
RT "Identification of the disulfide bonds in human plasma protein  
SP-40,40 (apolipoprotein-J).";  
RL J. Biochem. 112:557-561(1992).  
RN [16]  
RP CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=97477902; PubMed=9336835;  
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,  
RL Carr S.A., Crabb J.W.;  
RT "Identification and characterization of glycosylation sites in human  
serum clusterin.";  
RL Protein Sci. 6:2120-2133(1997).  
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH (APOPTOSIS).  
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; X14723; CAA32847.1; -;  
DR EMBL; M25915; AAA35692.1; ALT\_INIT.  
DR EMBL; J02908; AAA31765.1; ALT\_INIT.  
DR EMBL; M74816; AAA60321.1; -;  
DR EMBL; M64722; AAB06508.1; -;  
DR EMBL; L00974; AAA60567.1; -;  
DR EMBL; AF311103; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M63379; AAB06507.1; -;  
DR EMBL; M63376; AAB06507.1; JOINED.  
DR EMBL; M63377; AAB06507.1; JOINED.  
DR EMBL; M63378; AAB06507.1; JOINED.  
DR EMBL; A21577; CAA01560.1; -;  
DR PIR; A35833; A35833.  
DR PIR; A37816; A37816.  
DR PIR; B37816; B37816.  
DR PIR; A41386; A41386.  
DR PIR; PLO135; PLO135.  
DR PIR; PLO136; PLO136.  
DR PIR; S04662; S04662.  
DR PIR; S07433; S07433.  
DR PIR; S09339; S09339.  
DR PIR; S34056; S34056.  
DR PIR; S43646; S43646.  
DR SWISS-2DPAGE; P10909; HUMAN.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;  
KW Signal.  
FT SIGNAL 1 22  
FT CHAIN 23 449  
FT CHAIN 23 227  
FT CHAIN 228 449  
FT CHAIN 102 313  
FT DISULFID 113 305  
FT DISULFID 116 302  
FT DISULFID 121 295  
FT DISULFID 129 285  
FT CARBOHYD 86 86  
FT CARBOHYD 103 103  
FT CARBOHYD 145 145  
FT CARBOHYD 291 291  
FT CARBOHYD 354 354  
FT CARBOHYD 374 374  
FT CARBOHYD 28 28  
FT CONFLICT 47 47  
FT CONFLICT 52 52  
FT CONFLICT 305 305  
C -> S (IN REF. 9 AND 10).  
Q -> H (IN REF. 11).  
G -> O (IN REF. 11).  
C -> M (IN REF. 10).



```

Query Match      17.2%   Score 434;  DB 1;  Length 449;
Best Local Similarity 25.8%;  Pred. No. 2.2e-19;
Matches 124;  Conservative 96;  Mismatches 208;  Indels 52;  Gaps 14;

QY 12 MKPELLVIVCLLWKSHCAPTWK-----DKTAISENLKSFSEGEIDAEDEEKKALT 65
DB 1 MMKTLFLVGLL-----TWESQVLGDQTVSDNELQEMSGKYVKNKEIQNAV 51
QY 66 GIKOMKIMRKEKHTNLMSTLKKCKEEOEALKLLNEVQHEERLCRESLADSWG 125
DB 52 GVKQIKTLIEKTEERTLJLNLEAKKEDDALNETRESETKLKLPGVCNETMMALWE 111
QY 126 ECRSCLNNCMRIYT-TCQPSWSSVKNKIERFFKIKYQFLFPFHEDNEKDLPISEKLEK 184
DB 112 ECKPCLQKTCMKFARVCRSGSLVGRQLEFLNQSPPFYFWNGDR-----IDSLLEN 165
QY 185 DAQLTOMEDV-----FSQLTVDVNSLFRNFRVFMQOQEDQTFQSHFIS-DTDLTPEYF 239
DB 166 DRQTHMLDVNQDHFSSRASSIIDLFQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218
QY 240 FPAFSKEPMTKADLE-OCWDIPNPFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAP 298
DB 219 F--FPKSRIVRSLMPSPFPEPLFHFANFPQFLEIHEA-----QQAMDIFHFSAPF 267
QY 299 DHGLLSKMLPGQDGLGELDONLSRCFKPHEKCKQCAHLSEDC----PDVPAHLTEL 354
DB 268 QHPPTPEIREGDDRTVCREIRHNSTGCLRMKQCDCKREILSVDCSTNNPSQAKLRREL 327
QY 355 DEAIRLVNSQOYGQIOLMTKRLHEDTAVLVEKMRGFCGWSELANOAPETELIINSIQ 414
DB 328 DESIQVAERTRYINELLKSYQWMLNTSLLEQNLQNFQWVNSRLANLTQGEDQYILRVT 387
QY 415 VVPRIEHGNISKODETMMTDLISLPNSFTLKIPLESABESSNFIGYVVAKALQHF-KEH 473
DB 388 TVAS-HTSDSDVSGVTEVVVKLFDSQDPIVTVVPEVSRKNPKFEMETVAEKALQETRK 446

RESULT 2
CLUS_COTJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89239492; PubMed=2541393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
RT induced by different retroviral oncogenes in quail neuroretinal
RT cells.";
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
RT alternative promoters with distinct regulatory elements.";
RL Eur. J. Biochem. 229:215-223(1995).
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER

```

```

(BY SIMILARITY)
-!- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.
-!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
CC EMBL: X15825; CAA33823.1; -
CC EMBL: X80760; CAA56733.1; -
CC PIR: S07714; S07714.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Clu; 1.
CC SMART: SM00030; Clu; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
CC Glycoprotein; Signal.
-----
DR SIGNAL 1 18 POTENTIAL.
DR CHAIN 19 451 CLUSTERIN.
DR CHAIN 19 451 BETA-CHAIN (SMALL SUBUNIT) (POTENTIAL).
DR CHAIN 231 451 ALPHA-CHAIN (LARGE SUBUNIT) (POTENTIAL).
DR DISULFID 98 314 INTERCHAIN (BY SIMILARITY).
DR DISULFID 109 306 INTERCHAIN (BY SIMILARITY).
DR DISULFID 112 303 INTERCHAIN (BY SIMILARITY).
DR DISULFID 117 296 INTERCHAIN (BY SIMILARITY).
DR DISULFID 125 286 INTERCHAIN (BY SIMILARITY).
DR CARBOHYD 99 99 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CARBOHYD 141 141 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CARBOHYD 278 278 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CARBOHYD 355 355 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CARBOHYD 375 375 N-LINKED (GLCNAC. . .) (POTENTIAL).
DR CONFLICT 437 437 A -> P (IN REF. 1).
DR CONFLICT 445 451 KONNTIE -> SRTTP (IN REF. 1).
SQ SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;

Query Match      16.5%   Score 418;  DB 1;  Length 451;
Best Local Similarity 25.7%;  Pred. No. 2.1e-18;
Matches 118;  Conservative 94;  Mismatches 179;  Indels 68;  Gaps 15;

QY 45 LKSFSEGEIDAEVKKALTGKIKQKIMMERKEKHTNLMSTLKKCREBKQALKLNE 104
DB 27 LKOLSAAGSKYIDAEVENAINGVKMKTLMKDTSKHQAMLHTLETKKKKEAVKALE 86
QY 105 VQHELEERLCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVKNKIERFFKIQF 163
DB 87 KEKQLAEKQEVNCTMLSLWECKPCLKHTCMRVYSKMGCHSGSLVGRQLEELNRSPP 146
QY 164 LPFFHEDNEKDLPISEKLEK-----DAQLTOMEDVFSQLTVDVNSLFRNFRVFMQOQ 218
DB 147 SIWNGERIDDLDRQQRERRFEDLEERFGLMEDVEDIFQDSTQLYGPAPFFR---- 202
QY 219 EFDQTFQSHFISDTDLTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FQLFC 267
DB 203 -----TPPFGGFEAF-VPPQVRVHL-----VPRRLRELHPFFQHPM 240
QY 268 NFSVSIYESVSETITKMLKAIEDLPKQDKAPDH--GGILSKMLP-QQDRGLCGELDONLS 324
DB 241 HGFHRLFPQLFEMTQHLMDG-----HGAWEHPLGSGFATESRNFSTRMVCREIRNSA 294
QY 325 RCFKPFHEKCKQCAHLSEDC----PDVPAHLTELDAIRLVNSQOYGQIOLMTKRLHLE 380
DB 295 GCLRMRDECEKREILAVDVCSDTDPVQSOLREQFEDALRLAERFTRRYDOLLSAFOAEM 354
QY 381 DTAYLVKMRGFGWVSELANOAP-----ETELIINSIQVVPRIHEGNISKODETMMTD 434
DB 355 NTSSLLDLQNLNRQFGWVSRGLNLTQGNDRGLQVTTVFESK---TPNLEDP--SAPADQVVT- 408

```



[illegible]



```

Db 233 ----NFHDFQPF-----YDHIHQQAQAMDAHLQ-----RTPYHFPTM-EFTENNDRTV 277
QY 317 GELDONLSRCFKHEKQKQAHLSDC-----PDVPAHTELEAIFLVNVNQYQOIL 372
Db 278 KEIRHNSGCLRWKQCKQCEILEVDCSASNPOTQLLRQOLNASOLAKEFSRLYDQLL 337
QY 373 QMTRKKHLEDYALYKMRKRGFGWVSELANAPETE-----IIFNSIQVVPRIHEGNISK 426
Db 338 QSYQOKMLNTSALLKOLNEQFTWVSOLANTQSDQHYLVFTVNSHNSDPSIPSG----- 393
QY 427 ODEMTMTDLSILPSSNFTKPIPLEESAESNFGYVVAKALQHF 471
Db 394 -----LTKVIVKLFNSPPIITVTPQEVSSPNFENVAEKALQOYR 433

RESULT 7
CLUS_PIG STANDARD; PRT; 446 AA.
AC Q29549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Clusterin precursor (Complement cytolysis inhibitor) (CLI).
GN CLU.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92184774; PubMed=1544909;
RA Diemer V., Hoyle M., Baglioni C., Millis A.J.;
RT "Expression of porcine complement cytolysis inhibitor mRNA in
RT cultured aortic smooth muscle cells. Changes during differentiation
RT in vitro."
RL J. Biol. Chem. 267:5257-5264(1992).
CC -1- FUNCTION: NOT YET CLARIFIED, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER; LOWER
CC LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M84639; AAA31013.1;
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; CLA; 1.
DR SMART: SM00030; CLB; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 446
FT CHAIN 29 227
FT CHAIN 228 446
FT CHAIN 102 312
FT DISULFID 113 304
FT DISULFID 116 301
FT DISULFID 121 294
FT DISULFID 129 284
FT CARBOHYD 86 86
FT CARBOHYD 103 103

```

```

FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 333 333 N-LINKED (GLCNAC. . .) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (POTENTIAL).
SQ SEQUENCE 446 AA; 51774 MW; B1D5B434B668E3AA CRC64;

Query Match 15.8%; Score 400.5; DB 1; Length 446;
Best Local Similarity 24.4%; Pred. No. 2.3e-17;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 16 LLVFIIVCLLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEEVKALTGK 68
Db 4 LLLLVGILL-----TWENGPVVLGDKAISDELQEMSTEGSKYVYVNEIKNALEK 54
QY 69 QMKIMMERKEHTNLMLSTLKKREBKQKALKLLNEVQEHLEEEERLCRESLADSGECR 128
Db 55 QIKTLIEQSNNEERKSLSSLEAKKKEDALNDTRDTETKLGSGQLCNETMMALWEECK 114
QY 129 SCLENNCRMYT--TCQPSWSSVKNKTERFKIYQFLFPFHEDNEKDLPISEKLIKDAQ 187
Db 115 PCLKQTCMKFYARVCRSGSLVGHQLEEFNLQSSPFYFWINGDR-----IDSLMENDRQ 168
QY 188 LTQMEDVFSQTLVDVNSLENRSFNVFRMQQEFQDQFQSHFIS----DTDLTEPY----- 238
Db 169 QSHVMDI-----MEDSFNRASNI-----MDELQDRFFNREPFQTFQFFSGSSHR 214
QY 239 ---FFPAFSKEPMKADLEQCWDIPNFQFCNFSIYESVSETITKMLKALDEPKQD 295
Db 215 GSLFFNPKSRFARNIMPFLFTDL-NYHMFQPF---FDMIHQAQQAQMDAHLHRIYH- 268
QY 296 KAPDHGGLISKMLP--GQDRGLCGELDONLSRCFKHEKQKQAHLSDC-----PDVPA 349
Db 269 -PPEAG-----VPENSNDRAVCKEIRHNSGTCLRMKDQCEKREILSSASNSQMQ 321
QY 350 LHTDELALRVNVNQYQOILQMTKRLKLEDYALYKMRGFGWVSELANAPETEII 409
Db 322 LRQELYSLOMAEKFSKLYDQLQSYQOKMLNTSSLLKQLNEQFSWVSQLANLTQNDYR 391
QY 410 FNSIQVVPRIHEGNISKQDETMMTDLISLPSSNFTKPIPLEESAESNFGYVVAKALQH 469
Db 382 YLQVTVV-NSHGSDPSVPSGLTKVYVVKLFDSYPTILLIPEQVS--DPKFMETVAEALQ 438
QY 470 EKE 472
Db 439 YRQ 441

RESULT 8
CLUS_RAT STANDARD; PRT; 447 AA.
AC P05371;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid
DE glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM-
DE 2).
GN CLU.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.
RX MEDLINE=88000523; PubMed=3651384;
RA Collard M.W., Griswold M.D.;
RT "Biosynthesis and molecular cloning of sulfated glycoprotein 2
RT secreted by rat Sertoli cells."
RL Biochemistry 26:3297-3303(1987).
RN [2]
RP SEQUENCE FROM N.A.

```





```
QY 418 RIHE-GNISKQDETMMTDLSTLPSSNFTKLPLESAESSNFIQYVVAKALQHF 471
ID C1K1_YEAST STANDARD: PRT; 594 AA.
Db 1260 LMHEYSQLGKTFEDEKRALIASRONEELR-SLKSELESKRKLEVEYQKVLVEEV 1313

RESULT 11
C1K1_YEAST
AC Q01649;
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE Spindle POLE body associated protein.
GN C1K1 OR YMR198W OR YMR946.11.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetales; Saccharomycetaceae; Saccharomycetes.
OX NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92354905; PubMed=1644287;
RA Page B.D., Snyder M.;
RT "C1K1: a developmentally regulated spindle pole body-associated
RT protein important for microtubule functions in Saccharomycetes
RT cerevisiae.";
RL Genes Dev. 6:1414-1429(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Pearson D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: A DEVELOPMENTALLY REGULATED PROTEIN IMPORTANT FOR
CC MICROTUBULE FUNCTIONS. TIGHTLY ASSOCIATED WITH KAR3, MAY SERVE TO
CC REGULATE THE CELLULAR COMPARTMENT IN WHICH KAR3 FUNCTIONS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96439; AAA34494.1; -
DR EMBL; Z47815; CAA87820.1; -
DR PIR; A44073; A44073.
DR SGD; S0004811; C1K1.
KW Microtubules.
SQ SEQUENCE 594 AA; 69069 MW; A2A09DEB8C2EC838 GRC64;

Query Match
Best Local Similarity 5.8%; Score 145.5; DB 1; Length 594;
Matches 84; Conservative 71; Mismatches 138; Indels 109; Gaps 17;

QY 26 LKDSHCAPTWKDTAISENLKSFSEVG-----BIDAEVKKR----- 63
Db 147 LKDNQCDLQRKKNVELSSKNIVSMHAKYQVEFENDELEELSNAKREWTYKLMVEVNLKPDE 206
QY 64 -LTG-IKQWKIMWE---RK-----EKEHTNLMSTLTKCKREEKQAL-KLLNEVQEHLE-EE 112
Db 207 RLTDENRQLKTEFEVNRKLFLOQENENECKNYKELDKKFEIPKVKVNDARIELDGEQ 266
QY 113 ERLCR--ESLADSWGECSRLNNCMRIYTTQPSWSSVKNKIERFFRKIYQFLFPFHED 170
Db 267 ERLSKVLKDLQDTHGELKENIK-----TC-----RDEFNDFEKRIGEAEVNFHSM 311
QY 171 NEKDLPISEKLEKDAQLTQMEDVFSQLTVDVNSLNFNSFVFMQMQEFDQTFOSHPIIS 230
Db 312 ELAVVPLKKKLASTQALTVQVEERKQVGEAEANNKKIYVNELEKVVQBEL----- 361
QY 231 DTDLTPEYFFPAFSEKPEMTKADLE-----QCWDIPNFFOLFNCFSVS-----IYESVSET 280
```

```
Db 362 -----YTRQNLATSEIEIKGYTRCFAYANERQMPDEFHINXYVDRICENSNGS- 408
QY 281 ITKMLKATEDLPKODKAPDHGLISKPLPGDGRGLGCLGDLQNLSCFKFHEKCKQCAHL 340
Db 409 --KRQVQEDRVVLEIEHKDKRLYNECIPFLE-----KYISKLNCSIIIV 451
QY 341 SEDCPDVPALHTELDEAIRLVNSNQYQGILOQMTRK--HLE 380
Db 452 VSQOPTAPMKKTLKQLI-----EQYGENYKMTLNLHL 486

RESULT 12
MYHD_HUMAN
ID Q9UKX3; 095252; STANDARD: PRT; 1938 AA.
AC 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin heavy chain, skeletal muscle, extraocular (MyHC-ec).
GN MYH13.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Extraocular muscle;
RX MEDLINE=99318869; PubMed=10388558;
RA Weiss A., Schiaffino S., Leinwand L.A.;
RT "Comparative sequence analysis of the complete human sarcomeric myosin
RT heavy chain family: implications for functional diversity.";
RL J. Mol. Biol. 290:61-75(1999).
RN [2]
RP SEQUENCE OF 1917-1938 FROM N.A.
RC TISSUE=Extraocular muscle;
RX MEDLINE=99026150; PubMed=9806854;
RA Winters L.M., Briggs M.M., Schachat F.;
RT "The human extraocular muscle myosin heavy chain gene (MYH13) maps to
RT the cluster of fast and developmental myosin genes on chromosome 17.";
RL Genomics 54:188-189(1998).
CC -!- FUNCTION: MUSCLE CONTRACTION.
CC -!- SUBUNIT: MUSCLE MYOSIN IS A HEXAMERIC PROTEIN THAT CONSISTS OF 2
CC HEAVY CHAIN SUBUNITS (MHC), 2 ALKALI LIGHT CHAIN SUBUNITS (MLC)
CC AND 2 REGULATORY LIGHT CHAIN SUBUNITS (MLC-2).
CC -!- SUBCELLULAR LOCATION: Thick filaments of the myofibrils.
CC -!- DOMAIN: THE RODLIKE TAIL SEQUENCE IS HIGHLY REPETITIVE, SHOWING
CC CYCLES OF A 28-RESIDUE REPEAT PATTERN COMPOSED OF 4 HEPTAPEPTIDES,
CC CHARACTERISTIC FOR ALPHA-HELICAL COILED COILS.
CC -!- PTM: TWO CYSTEINE RESIDUES IN THE S1 DOMAIN ARE SELECTIVELY
CC ALKYLATED AND ARE REQUIRED FOR MYOSIN ATPASE ACTIVITY.
CC -!- MISCELLANEOUS: EACH MYOSIN HEAVY CHAIN CAN BE SPLIT INTO 1 LIGHT
CC MEROMYOSIN (LM) AND 1 HEAVY MEROMYOSIN (HM). IT CAN LATER BE
CC SPLIT FURTHER INTO 2 GLOBULAR SUBFRAGMENTS (S1) AND 1 ROD-SHAPED
CC SUBFRAGMENT (S2).
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF111782; AAD29948.1; -
DR EMBL; AF075248; AAC83241.1; -
DR HSSP; P08799; IMND.
DR MIM; 603487; -
DR InterPro; IPR000048; IQ.
DR InterPro; IPR004009; Myosin_N.
DR InterPro; IPR002928; Myosin_tail.
```



```

DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00663; myosin_head; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR Pfam: PF01576; Myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR PRODOM: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS50096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ.
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.88; Score 145.5; DB 1; Length 1938;
Best Local Similarity 20.7%; Pred. No. 0.27;
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 41 ISEMLKSFVGEIDAEV-KKALTIKIMMERKEHEHTNLMSTLKKCR-----EKK 95
DB 1150 ISEMLEAS--GATSAQIENKKREAFQKRRDLDEATLQHEATATLRKQADSVAE 1207
QY 96 QEALKLLNEVOEHLLEERLCRESLAD--SWGECRSCLENNCMRIYTTQPSWSSVKNKI 153
DB 1208 GEQIDNLRQVKQLEKEKSELMKMEIDDMASNEALSXSNIERTCTVEDQFSEIKAD 1267
QY 154 ERFRKIYQFLFPFHEDNEKDLPISEKLTEKDAQLTQMEDVFSQLVVDVNSLFRNFSNVF 213
DB 1268 EQQTOLI-----HDLNMOKARLQTONGELSHRVEEKESLTSQLTQKQALTOQLEELK 1320
QY 214 ROMQOQFD-QTFQSHFIS-----DTDLTEPYFFPAFKEPMTKADLRBOCV-----DIPNFF 263
DB 1321 ROMEETKAKNAHAHALQSSRHDCDLLR---EQEEQEAQKAEQALQALSKANSEVAQWK 1376
QY 264 QLCNFSVSVIYESVSETITKMLKAIEDL-PKQKADPHGGLISK-----MLPGQDRGLGCEL 319
DB 1377 TKYETDAIQRTTELEEAQKLAQRLQEAEBEKTETANSKASLEKTKORLOGEVEDLMRDL 1436
QY 320 QDMLSRC-----PFHEKCKQCAHLSEDCPDVPALHTELDEAIRLVNVS 365
DB 1437 ERSHTACATLDKQRFNFKVLAEWKQKLDSEQAEEAQAQESRSUSTEL---FKMRNAYE 1493
QY 366 QYQGI--IQMTKKHLEDYALVLEKMRGFGWVSELANQAPET-----ELI 409
DB 1494 EYVDQLETLRRENKNLQEE-----ISDLTEQIAETGNLQEAETKKLIVEQE 1540
QY 410 ENSIQVVPRIHEGINSQDETMM---TDLISLPSSNFTLIKPLESAESSNFIQGVVAKA 466
DB 1541 KSDIQVALEVEGSLHEEHSKILRVOLLSQVKSLEDRKVIKDEIEQLKRNQRAAEA 1600
QY 467 LQ 468
DB 1601 LQ 1602

RESULT 13
ID RA50_METJA
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2561 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii".
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mre11 complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mre11 by unwinding
CC and/or repositioning DNA ends into the mre11 active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mre11 (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U67572; AAB99331.1;
CC TIGR: MJ1322.
DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_C; 1.
DR Pfam: PF02463; SMC_N; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9B8B48173E788F3 CRC64;

Query Match 5.68; Score 141.5; DB 1; Length 1005;
Best Local Similarity 20.0%; Pred. No. 0.22;
Matches 81; Conservative 70; Mismatches 131; Indels 123; Gaps 16;

QY 36 KDKTAISENL-----KSFSEGEIDAEVKKALTGIKOMTKMERKEHEHTNLMST 87
DB 419 QEKKSIEKNINDLEPRINKLLEETKNIDI-ESIENSLKEIEKKVLENQKELNKK 477
QY 88 LKCKREEQKALKLLNEVOEHLLEERLCRESLADSWGECRSL-----ENNCRIYITTCQ 143
DB 478 LGENSE----IKRLKKILDELKEVE-----GKCPCKTPTIDENKKMELINQHK 522
QY 144 PSWSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIEKDAQLTQMEDVFSQLVVDVN 203
DB 523 TQLNNKYTELEINKKIREI-----EKDIEKLKKEIDKEENLKTLLYLEKQSQIE 574
QY 204 SLFNRSFNVFRMQOQEFDTQFSHFISDTDLTEPYFFPAFSEKPMTKADLRQCDWIPNFF 263
DB 575 EELKLLKN-YKEQLDEINKISNVYNGKPYDE-----ILE---DIKSQL 615
QY 264 QLFCNF-----SVSIYESVSET-ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLC 316

```



Db 1813 LQGEEMLHOPMEDN-----KKEKIRLQLL-----LLHTRYNK-IKAIPQORK 1855

QY 368 YGQILQMTKRLHEDTAYLVKMRGQFGWSELANQAPETELIENSIVQVPRHEGNLSKO 427

Db 1856 MGQLASIRSLPTDYLVE-----INKILLCDDVLSLNPVLPETAIYE-DFSQ 1906

QY 428 DETM 431

Db 1907 EDSL 1910

RESULT 15

CENE\_HUMAN STANDARD; PRT; 2663 AA.

AC Q02224;

DT 01-JUL-1993 (Rel. 26, Created)

DT 01-JUL-1993 (Rel. 26, Last sequence update)

DT 30-MAY-2000 (Rel. 39, Last annotation update)

DE Centromeric protein E (CENP-E protein).

GN CENPE.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RX SEQUENCE FROM N.A.

RX MEDLINE=93024922; PubMed=1406971;

RA Yen T.J., Li G., Schaar B.T., Szilak L., Cleveland D.W.;

RT "CENP-E is a putative kinetochore motor that accumulates just before

RT mitosis.";

RL Nature 359:536-539(1992).

RN [2]

RX CHARACTERIZATION.

RX MEDLINE=95196755; PubMed=7889940;

RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;

RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed

RT microtubule motor.";

RL EMBO J. 14:918-926(1995).

RN [3]

RX CHARACTERIZATION.

RX MEDLINE=98437347; PubMed=9763420;

RA Chan G.K.T., Schaar B.T., Yen T.J.;

RT "Characterization of the kinetochore binding domain of CENP-E reveals

RT interactions with the kinetochore proteins CENP-F and hBUBR1.";

RL J. Cell Biol. 143:49-63(1998).

CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE

CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE

CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT

CC AND/OR SPINDLE ELONGATION.

CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.

CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING

CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS

CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.

CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).

CC -----

DR EMBL; Z15005; CAA78727.1; -

DR PIR; S28261; S28261.

DR HSSP; P17119; 3KAR.

DR MIM; 117143; -

DR InterPro: IPR001752; kinesin.

DR Pfam; PF00225; kinesin; 1.

DR PRINTS; PR00380; KINESINHEAVY.

DR SMART; SM00129; KISC; 1.

DR PROSITE; PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.

DR PROSITE; PS50067; KINESIN\_MOTOR\_DOMAIN2; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;

KW Cell cycle; Centromere.

FT DOMAIN 1 335 KINESIN-MOTOR.

FT DOMAIN 336 2471 COILED COIL (POTENTIAL).

FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).

FT NP\_BIND 86 93 ATP (BY SIMILARITY).

SQ SEQUENCE 2663 AA; 312087 MW; CEFC13880C8C8B8 CRC64;

Query Match 5.5%; Score 139.5; DB 1; Length 2663;

Best Local Similarity 18.1%; Pred. No. 0.88;

Matches 100; Conservative 87; Mismatches 175; Indels 191; Gaps 23;

QY 41 ISENKASFSEVGEIDADEVKKALTGK-----QMKIMMERKEKEHNLMSTLK----- 89

Db 1461 LKENKEI-VAKHLETEELKVAHCLKEQEBETINELVNLSEKETEISTIQKLEAIND 1519

QY 90 ----KCRE--EKQEA--KLLNEVQSHLE-----EERLCRESLADSWGECRSCLENNCM 136

Db 1520 KLONKIQEIYKEEQNLNIQISEVQENVNELKQFKEHRAKDSALQS-----IESKML 1572

QY 137 RYTTCPQSSSVK-----NRIERFFPKYQFL- 164

Db 1573 ELTNRLQESQEEIQIMKEEMKRVQEAQIERDQKENTKEIVAKMKESQEKYQFLK 1632

QY 165 -PPFHEDNEKDLPISEKLEKDAQLTOMEDVFSQLTVDVNSLNFNSFVFRQMQEFDQT 223

Db 1633 MTAVNETQKMCIEHLEKQFQKLENLEIETE-NIRLTQILHENLEEMRSVTKERD-- 1689

QY 224 FQSHFISDLDLTPYFPFAPFSKEPMTKADLEQCWDIPNFFQFQFCNFSVSIYESVSTIK 283

Db 1690 -----DLR-----SVEETLKVERDQ-----LKENLRETIYR 1715

QY 284 MLKAIEDLPKQDK-----APDHGLISKMLPGQDRGLCGELDQNLSC----- 326

Db 1716 -----DLEKQELKIVHMHLEHETIDKL-----RGIVSEKTNISNMQKDLHSNDA 1764

QY 327 -----FKFHEKQKQCAHLSDCPDVPAL-----HTELDEAIRL 360

Db 1765 LKAQDLKIQEELRIAHMHLEKQOETDKLRGIVSEKTDKLSNMQKLENSNAKLEKIOE 1824

QY 361 VVNSNQYGOILQMTKRLHEDTAYLVKMRGQFGWSELANQAPETELIENSIVQV----- 415

Db 1825 LKANEHQ-----LITLKKDVNETQKKVSEM-----EQLKKQIKDQSLTSLKLEIENL 1873

QY 416 VPRHEG-----NISKQDETMTDLSILPSSNFTLKIPLEESAESNFTGYV 463

Db 1874 AOELHENLEEMKSVKERNDRNRVEETLKERD-----OLKESLQETKARDLEIQOEL 1926

QY 464 AKALQHFKEHFKT 476

Db 1927 KTARMLSKHEKET 1939

Search completed: July 2, 2002, 11:58:16

Job time: 369 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:57:10 ; Search time 53.04 Seconds  
(without alignments)  
1555.780 Million cell updates/sec

Title: US-09-722-544A-4MOD  
Perfect score: 2529.  
Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

- 1: sp\_archaea:\*
- 2: sp\_bacteria:\*
- 3: sp\_fungi:\*
- 4: sp\_human:\*
- 5: sp\_invertebrate:\*
- 6: sp\_mammal:\*
- 7: sp\_mhc:\*
- 8: sp\_organelle:\*
- 9: sp\_phage:\*
- 10: sp\_plant:\*
- 11: sp\_rodent:\*
- 12: sp\_virus:\*
- 13: sp\_vertebrate:\*
- 14: sp\_unclassified:\*
- 15: sp\_rvirus:\*
- 16: sp\_bacteriap:\*
- 17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2460	97.3	466	4 Q15846	Q15846 homo sapien
2	1832.5	72.5	465	6 Q93KN1	Q95knl canis famli
3	1226.5	48.5	338	6 Q9NIT8	Q9nit8 canis famli
4	423	16.7	449	6 Q29482	Q29482 equus cabal
5	402.5	15.9	448	13 Q9YGP0	Q9ygp0 gallus gall
6	321	12.7	372	11 Q9JK98	Q9jk98 mus musculu
7	266.5	10.5	295	11 Q35510	Q35510 rattus norv
8	204	8.1	218	11 Q76329	Q76329 dictyosteli
9	164	6.5	1738	5 Q9W6V0	Q9w6v0 gallus gall
10	161	6.4	3616	13 Q9W6V0	Q9w6v0 gallus gall
11	150	5.9	1156	16 Q66878	Q66878 aquifex aeo
12	149.5	5.9	840	6 Q95JTR0	Q95jtr0 macaca fasc
13	149	5.9	1388	6 Q28021	Q28021 bos taurus
14	148.5	5.9	1091	5 Q18082	Q18082 caenorhabdi
15	147	5.8	3259	4 Q14789	Q14789 homo sapien
16	145	5.7	1388	4 Q75116	Q75116 homo sapien

17	145	5.7	1388	4 Q9UQNS	Q9uqn5 homo sapien
18	144	5.7	1379	11 Q62868	Q62868 rattus norv
19	143	5.7	1931	5 Q9NCF9	Q9ncf9 drosophila
20	143	5.7	1931	5 Q9VKH9	Q9vkh9 drosophila
21	142.5	5.6	533	6 Q95JY2	Q95jy2 macaca fasc
22	140.5	5.6	2166	16 Q51465	Q51465 borrelia bu
23	140	5.5	1132	4 Q75065	Q75065 homo sapien
24	140	5.5	1305	10 Q9FJ35	Q9fj35 arabidopsis
25	138.5	5.5	1391	11 Q922J3	Q922j3 mus musculu
26	138	5.5	1084	16 Q83423	Q83423 treponema p
27	138	5.5	1955	5 Q61308	Q61308 parascaris
28	136.5	5.4	684	3 Q07238	Q07238 pneumocysti
29	136.5	5.4	709	5 Q9GRG1	Q9grg1 tetrahymena
30	136	5.4	1388	11 P70336	P70336 mus musculu
31	135.5	5.4	1956	5 Q20641	Q20641 caenorhabdi
32	135	5.3	1133	5 Q21022	Q21022 caenorhabdi
33	135	5.3	1330	6 Q97961	Q97961 vulpes vulp
34	135	5.3	1530	4 Q43241	Q43241 homo sapien
35	135	5.3	5458	5 Q9U459	Q9u459 plasmodium
36	134.5	5.3	1300	4 Q13999	Q13999 homo sapien
37	134.5	5.3	2035	11 Q9JMH9	Q9jmh9 mus musculu
38	134	5.3	1538	13 Q98940	Q98940 gallus gall
39	133.5	5.3	720	4 Q9H6Q7	Q9h6q7 homo sapien
40	133.5	5.3	1410	4 Q14221	Q14221 homo sapien
41	133.5	5.3	2138	5 Q9XZE3	Q9xze3 amoeba prot
42	133	5.3	1938	6 Q9GJP9	Q9gjp9 oryctolagus
43	132.5	5.2	1708	5 Q9U0S6	Q9u0s6 mytilus gal
44	132.5	5.2	2473	11 Q9QZ84	Q9qz84 mus musculu
45	132	5.2	1033	4 Q9UFE1	Q9ufel homo sapien

#### ALIGNMENTS

RESULT 1

Q15846 PRELIMINARY; PRT; 466 AA.  
AC Q15846;  
DT 01-NOV-1996 (TREMBLrel. 01, Created)  
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE-NEURORETINA;  
RA Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,  
RA Matsubara K.;  
RT "Expression profile of active genes in human retina.";  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D63813; BAA09882.1; +.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 3.  
DR SMART; SM00035; Cla; 1.  
DR SMART; SM00030; CLb; 1.  
KW Signal; Hypothetical protein.  
FT SIGNAL 1 20 POTENTIAL.  
FT CHAIN 21 466. ROD PHOTORECEPTOR PROTEIN.  
SQ SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;

Query Match 97.3%; Score 2460; DB 4; Length 466;  
Best Local Similarity 99.8%; Pred. No. 2.2e-160;  
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 12 MKPPLVFIICLLWLKDSHCAPTWKDKTALSENLSKFSFSEVGEIDAEVVKALTGKQMK 71

DB 1 MKPPLVFIICLLWLKDSHCAPTWKDKTALSENLSKFSFSEVGEIDAEVVKALTGKQMK 60

QY 72 IMMEREKEHTNLMSTLKKCRKQKALKLLNEVQHEERLCRESLADSNQECRSCL 131

```

Db 121 ENNCRIYTTCPQSWSSVKNKIERFFRKIYQFLPFPHEDNEKOLPISEKLIETEDAQLTQM 180
Qy 192 EDVFSQLTVDVNSLNFNRSENFVFMQOEFDQTFQSHFISDLDTEPEYFFPAFSEKPTKA 251
Db 181 EDVFSQLTVDVNSLNFNRSENFVFMQOEFDQTFQSHFISDLDTEPEYFFPAFSEKPTKA 240
Qy 252 DLEQWCDIPNFFQFQFNFVSIVSESVSTIITKMLKATIEDLPKQKAPDHPGGLISKMLPGQ 311
Db 241 DLEQWCDIPNFFQFQFNFVSIVSESVSTIITKMLKATIEDLPKQKAPDHPGGLISKMLPGQ 300
Qy 312 DRGLCGELDQNLSCRCFKFHEKQRCQAHLSDEDCPDVPALHTELDIAIRLVNVSNQYQGI 371
Db 301 DRGLCGELDQNLSCRCFKFHEKQRCQAHLSDEDCPDVPALHTELDIAIRLVNVSNQYQGI 360
Qy 372 LQMTKRKHELDIAYLVKMRGQFGWVSELANQAPETEIIFNSIQVVPRIHEGNIKSQDETM 431
Db 361 LQMTKRKHELDIAYLVKMRGQFGWVSELANQAPETEIIFNSIQVVPRIHEGNIKSQDETM 420
Qy 432 MTDLISLPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKEHFKTW 477
Db 421 MTDLISLPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKEHFKTW 466

RESULT 2
Q95KN1 PRELIMINARY; PRT; 465 AA.
AC Q95KN1;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
OS Canis familiaris (Dog).
OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA;
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF241221; AAK49030.1; -.
DR SMART; AF241221; AAK49030.1; -.
SQ SEQUENCE 465 AA; 54402 MW; BBDEIAC512D5D33F CRC64;

Query Match 72.5%; Score 1832.5; DB 6; Length 465;
Best Local Similarity 73.2%; Pred. No. 1.7e-117;
Matches 341; Conservative 52; Mismatches 72; Indels 1; Gaps 1;

Qy 12 MKPDLVFIYVLLWLDKSHCAPTHKDKTATSENKLSSEVGEIDADDEVKALTGIKQMK 71
Db 1 MKPSLLVFTVLLWKDCHCAPTHKDKTATDMHGNLKGFESEAGDIDYDEVKALTGIKQMK. 60
Qy 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEQHELEBEERLCRESLADSWGECRSCL 131
Db 61 IMMERREEHTNLMSTLKKCKREEKQKALKLNVEQHELEBEESLCQVSLTDSWDECSCL 120
Qy 132 ENNCRIYTTCPQSWSSVKNKIERFFRKIYQFLPFPHEDNEKOLPISEKLIETEDAQLTQM 191
Db 121 ESNCRFHTTTCQPSWSSVKNKIEVFNNIYQFLPFPHEDNEKOLPISEKLIETEDAQVAQI 180
Qy 192 EDVFSQLTVDVNSLNFNRSENFVFMQOEFDQTFQSHFISDLDTEPEYFFPAFSEKPTKA 251
Db 181 ENFNQLTVDVNFNFNRSLNVFKMQOEFDQTFQSHFISDLDTEPEYFFPAFSEKPRKKA 240
Qy 252 DLEQWCDIPNFFQFQFNFVSIVSESVSTIITKMLKATIEDLPKQKAPDHPGGLISKMLPGQ 311
```

```

Db 241 DPVQSWDIPSEFFQFYNFYSLSIYHSITITTKLTNAIEDLPKQDNDNSHGSLSKTLPVQ 300
Qy 312 DRGLCGELDQNLSCRCFKFHEKQRCQAHLSDEDCPDVPALHTELDIAIRLVNVSNQYQGI 371
Db 301 HRGYPGEGQNLSECFQFHARQKQCDYLWEDCPDVELHTKVDEALELVNISHQQAQV 360
Qy 372 LQMTKRKHELDIAYLVKMRGQFGWVSELANQAPETEIIFNSIQVVPRIHEGNIKSQDETM 431
Db 361 LQMTQHLEDDTYLMEKMRERFEGWADLANQAPCAENIFDSTKMPNTHIEGNEFSKQDETM 420
Qy 432 MTDLISLPSSNFTLKIPLEESAESSNFTGYVYVAKALQHFKEHFKTW 477
Db 421 I-DUSILSSPNFTLKIPLEESAETSNTFISYMLEKAVQHFKHFKTW 465

RESULT 3
Q9N1T8 PRELIMINARY; PRT; 338 AA.
AC Q9N1T8;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF147784; AAF36799.1; -.
DR InterPro; IPR000753; Clusterin.
DR SMART; SM00035; CLA; 1.
DR NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF147784; AAF36799.1; -.
DR InterPro; IPR000753; Clusterin.
DR SMART; SM00035; CLA; 1.
DR NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.; B86F07877213FCC2 CRC64;

Query Match 48.5%; Score 1226.5; DB 6; Length 338;
Best Local Similarity 69.8%; Pred. No. 3e-76;
Matches 231; Conservative 42; Mismatches 57; Indels 1; Gaps 1;

Qy 147 SSVKNKIEREFRKIYQFLPFPHEDNEKDLPISEKLIETEDAQVAQIENVFNQLTVDVRF 68
Db 9 TSGTSTVEQEFERNIYQFLPFPHEDNEKDLPISEKLIETEDAQVAQIENVFNQLTVDVRF 68
Qy 207 NRSFNVRMQQEFDDQTFQSHFISDLDTEPEYFFPAFSEKPTKADLEQCWDIPNFFQF 266
Db 69 NRSFNVRMQQEFDDQTFQSHFISDLDTEPEYFFPAFSEKPTKADPVQSWDIPSEFQF 128
Qy 267 CNRSVSIVSESVETITKMLKATIEDLPKQKAPDHPGGLISKMLPGQDRGLCGELDQNL 326
Db 129 YNFSLSIYHSITITTKLTNAIEDLPKQDNDNSHGSLSKTLPVQHRGYPGEQNLSEC 188
Qy 327 FKHEKQKQAHLSDEDCPDVPALHTELDIAIRLVNVSNQYQGIQMTKRKHELDIAYLV 386
Db 189 FQFHARQKQCDYLWEDCPDVELHTKVDEALELVNISHQQAQVQVQVQVQVQVQVQV 248
Qy 387 EKMRGQFGWVSELANQAPETEIIFNSIQVVPRIHEGNIKSQDETMMDLSILPSSNFTLK 446
Db 249 EKMRGQFGWVSELANQAPCAENIFDSTKMPNTHIEGNEFSKQDETM I-DLSILSSPNFTLK 307
Qy 447 IPLEESAESSNFTGYVYVAKALQHFKEHFKTW 477
Db 308 IPLEESAETSNTFISYMLEKAVQHFKHFKTW 338

RESULT 4
```

```
Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (TREMELREL. 01, Created)
DT 01-NOV-1996 (TREMELREL. 01, Last sequence update)
DT 01-DEC-2001 (TREMELREL. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OLIVER; TISSUP-TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
RT Clusterin";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; L46797; AAA80313.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; Cla; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
SQ SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;

Query Match 16.7%; Score 423; DB 6; Length 449;
Best Local Similarity 24.5%; Pred. No. 3.5e-21;
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps 13;

QY 16 LLVFIYICLLWLKDSHCAPTWKOKTAISENLKSFSEVGETDADEEVKALTGTQKIMME 75
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
4 LLLVGLLTLENGQ---VLGKAVSDRELQEMSTQGSNYINKEIKNALGVKQIKNLIE 60
QY 76 RKEKHTNLMLTKKREKQEQALKLLNEVQHEERLCRESLADSWGECRSCLENNC 135
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
61 QTNNEERKSLGLTLEBAKKKEGALNDTKDSEMKLQESQGVNCTWALWEECKPCLKQTC 120
QY 136 MRYIT-TCQPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQMEDV 194
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
121 MKFYARVCRSGSLVGHQLEELFNQSSPYFWINGDR-----IDSLLENDRQQTHTLDV 174
QY 195 FSQLTVDVNSLNRSNVFRQOEEDQTFQSHFISDTLTETPEYFFPAFSEKPEMTKADI- 253
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
175 -----MQDSFDRASSI-----MDLFDQRFET-REPQDTYYTSPFS-SPHRRSLL 218
QY 254 -----EQCWDP-----NFFQCFNFSVSIYESVETITKMLKAIEDLPKQD--KAP 298
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
219 FNPKSFAFNIMHFPYRHLNENDMFQPFDMHQQAQAMNHLHLPQLPMTWSESGD 278
QY 299 DHGGLIKMLPGDRLGCELQONLSRCPKFHEKQKQAHLSDEC-----PDVPALHTEL 354
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
279 NH-----DRTVCYKEIRHNSGTCLMKMDQCEKQCEILSDVDCSTNNPQOMQLRQEL 327
QY 355 DEAIRLVVNSQOYQOILQMTKKHLEDATYLVKMRGQFGWSELANOAPETEIFNSIQ 414
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
328 NNSLQAEFTKLYDELQSYQEKMLNTSSLKLNQEQFWSVSQLNTQGGEDQYLYQVT 387
QY 415 VVPRHIEGNISKODETMTDLSLPSSNFTLKIPEESAESSNFYGVVAKALQHFKE 472
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
388 TVSS-HNSDSEVPISGLTRVVVKLFDSYPTIVVPEVVSRRNPKFMTVAEKALQEYRQ 444
```

```
RESULT 5
QYGP0
ID QYGP0 PRELIMINARY; PRT; 448 AA.
AC QYGP0;
DT 01-MAY-1999 (TREMELREL. 10, Created)
DT 01-MAY-1999 (TREMELREL. 10, Last sequence update)
DT 01-DEC-2001 (TREMELREL. 19, Last annotation update)
DE CLUSTERIN.
OS Gallus gallus (Chicken).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;
OC Gallus.
OX NCBI_TaxID=9031;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE-99134326; PubMed-9933595;
RA Mahon M.G., Lindstedt K.A., Hermann M., Nimpf J., Schneider W.J.;
RT "Multiple involvement of clusterin in chicken ovarian follicle
RT development. Binding to two oocyte-specific members of the low density
RT lipoprotein receptor gene family.";
RL J. Biol. Chem. 274:4036-4044(1999).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR EMBL; AF119370; AAD17257.1; -.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; Cla; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Glycoprotein.
SQ SEQUENCE 448 AA; 51348 MW; 9982873DCAL3D27C CRC64;

Query Match 15.9%; Score 402.5; DB 13; Length 448;
Best Local Similarity 26.0%; Pred. No. 8.7e-20;
Matches 118; Conservative 87; Mismatches 189; Indels 59; Gaps 14;

QY 45 LKSFSEVGEIDAEVKKALTGKIMMERKEHTNLMLSTLKKCREKEQALKLINE 104
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
27 LKQLSAAASKYIDTEVENAINGVKQKMTLMDKTSKEHQAMLTLETKRKEEYKLALE 86
QY 105 VOEHLEERLCRESLADSWGECRSCLENNCMRYIT-TCQPSWSSVKNKIERFRKIYQF 163
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
87 KEKQLAEQVNCVNETMLSLWEECKPCLKHTCMRVYKICHSGSLVGHQLEELLNRSPPF 146
QY 164 LPPFHEDNEKDLPISEKLIK-----DAQLTQMEDVFSQLTVDVNSLNRSNVFRQMOQ 218
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
147 SIWNGERIDALLDREQRERFEDLEERFGLMEDGVEDIFQDSTQLYGPAPPPFR---- 202
QY 219 EFDQTFQSHFISDTLTETPY--FFPAFSEKPEMTKADLQCEWDIPNFFQL-----FCNFS 270
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
203 -----TPPFGGFREAF-VPPVQVRVL-----VPPRRRLSRELHPFLOHP 240
QY 271 VSIYESVSETITKMLKAIEDLPKQKADPH--GGILSKMLP-GQDRGLCGLDQNLRSFCF 327
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
241 VGHFRLFEMTQRLDGG-----GHGAWDHLLGGFSESRNEFSTDRMVCREIRRSAGCL 294
QY 328 FHEKQKQKQAHLSDEC-----PDVPALHTELDEATRLVNSNQYQOILQMTKKHLEDTA 383
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
295 RMRDCEKCRREILAYDSCQTDVPVQSQLREQFEDALRLAERFTRRYDDLLSAFAQAEMLNTS 354
QY 384 YLVEKMRGQFGWSELANOAPETEIFNSIOV---VPRHIEGNISKODETMTDLSILPS 440
Db : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
355 SLLDGLNRQFGVLRGLNLTQGTGFLQVTVFSTPNLEDP--SAPADTQVT-VQLFDS 411
```

QY 441 SNFTLKIPLEESAESSNFICYVVAKALQHKEH 473  
:| : | : | : | : ||||| :  
Db 412 EPLSLTPVGDISWDDPPREMETVAEQALQHYKON 444

[illegible]

Query Match	12.7%	Score 321;	DB 11;	Length 372;
Best Local Similarity	23.6%;	Pred. No. 2.6e-14;		
Matches	93; Conservative	83; Mismatches	170;	Indels 48; Gaps 12;
Qy	95	KOALKLLNVEQHELEBERLCRLADSWGECRSCLENNCMRLYT--TCQPSWSVSKNKI	153	
Db	4	REDALETRDSEMKLKAFPEVCNETHMALNEECKPLCKHTCKFKYARVCRSGSGLVGQOL	63	
Qy	154	ERFRKRYQFLFPFHEDNEKDLPISEKLIKDAOLTO----MEDVFSQLTVDVNSLNFNR	209	
Db	64	EELFNQSSPFYFWMGDR-----IDSLLEDROQSQVLDAMQDSFARASGIIDTLFQDR	117	
Qy	210	FNVRQMQOEDQTFQSHIFISDTLTPYFFP-----AFSKEPATKADLEQC-WDIPNF	262	
Db	118	F--FAR-----ELUHPHYFS-----PIGFPHKRPHPFLYPKSLRVLNSLMSPSHYGPSPF	163	
Qy	263	FQLCFNFSVSIVYESVSETITKMLKAIEDLPQDKAPDHGGLISKMLPGQ--DRGLCGELDQ	321	
Db	164	HNMQPFEMIHQAQAMQVQLHSPAQFPDQV-----FLREGEDDRTVCKEIRR	213	
Qy	322	NLSRCFFHEKQCKQAHLSEDC-----PDVPALHTELDEARLVNVSNQQYQIILWTRK	377	
Db	214	NSTGCLKMKGQCKCEQETLSYDCSTNNPAQANRLQELNDSLQVAERLQYKELLQSPQS	273	
Qy	378	HLEDTAYLVEKMRQFGVWSLILANQAPETETIENSIVQVPRIHGKINISKQETMMTDLSI	437	
Db	274	KMLNTSSLLEQLNDQFNWVSQLANTQGEDKYLYLRVSTV--TTHSSDSEVPSPRVTEVVVKL	332	
Qy	438	LPSSNFTLUKIPLESABSSNFIQYVYAKALQHFK	471	
Db	333	FDSOPITVLPPEEVS KONPKFMDTVAEKALOEYR	366	

RESULT	7
ID	Q35510 PRELIMINARY; PRT; 295 AA.
OC	A035510;
DT	01-JAN-1998 (TrEMBLrel. 05, Created)
DT	01-JAN-1998 (TrEMBLrel. 05, Last sequence update)
DT	01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE	CLUSTERIN.
DE	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX	NCBITaxID=10116;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=WISTAR; TISSUE=SEMINAL VESICLE;
RX	MEDLINE=98163873; PubMed=9503143;
RA	Izawa M.
RT	"Identification of a transcript predicting an alternative form of
RT	sulfated glycoprotein-2 (clusterin) in rat tissues.";
RL	Biochem. Mol. Biol. Int. 44:19-19(1998).
CC	-I- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC	OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC	AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC	CELL DEATH (BY SIMILARITY)
CC	-I- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC	SIMILARITY).
CC	-I- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
DR	EMBL; D11339; BAA21775.1; -.
DR	InterPro; IPR000753; Clusterin.
DR	Pfam; PF01093; Clusterin; 1.
DR	SMART; SM00035; Cla; 1.
DR	SMART; SM00030; Clb; 1.
DR	PROSITE; PS00492; CLUSTERIN_1; 1.
DR	PROSITE; PS00493; CLUSTERIN_2; 1.
KW	Glycoprotein.
SQ	SEQUENCE 295 AA; 34115 MW; C1E07FA745DD5470 CRC64;
Query Match	10.5%; Score 266.5; DB 11; Length 295;
Best Local Similarity	25.2%; Pred. No. 1e-10;
Matches	80; Conservative 59; Mismatches 131; Indels 47; Gaps
Qy	115 LCRESLADSGEGRCSLENMCRITYT--TCQPSWSYKVKIKIERFKIYFLPFPFHEDNEK 173 :   :    :    :    :    :    :    :    :    :    :    :    :    :    :
Dd	9 VCNETMALMECEKPCUKHTCKMFYARVCRGSGSLVGRLKEELFNSSPFYFMNGDR-- 66 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	174 DLPISEKLIEKDQLTQ---MEDVFSQLVDVNSFNRSFNVRMQQEFDDQTFSHF 229 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	67 ----IDSLESDRQSQVLAMQDSFRASSGIIDLTFQDFR--FTHEPQDI-----HHF- 114 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	230 SDTLTPFYFPF-----ATSKPEMTKADLEQCWDIP-NFFQLFCNFPSVIYESVSETTT 282 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	115 -----SPMGPPHKRHPLYPKSLRVLRSLPLSHYGPLSPHNNFPFFDMHQQAQAMDV 168 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	283 KMLKAIEDLPQDKAPDHGGGISKMLPQG--DRGLCGELDNLSCRFKFEKCKCQCAHLS 341 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	169 QLHSPAQLQFPDVDFLKE-----GEDDPYCKEIRHNSTGCCLKMKGQCEKCQEILS 218 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	342 EDC----PDVPALHTELDEAIRLVNSNQOYGTQLQMTRKHLEDTAFLVEKMRGQFGWS 397 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	219 VDCSTNNPAQANLRLQNDSLOVAERLTQQYNELLHSLSQKMLNTSSLQLNDQFSWS 278 : :   :   :   :   :   :   :   :   :   :   :   :   :   :   :
Qy	398 ELANOAPETEIFNISIQ 414 :     :   :   :   :   :   :   :   :   :   :   :   :   :
Dd	279 QLANLTGDDQYFGSPQ 295 :     :   :   :   :   :   :   :   :   :   :   :   :   :

RESULT	8	
Q9ERD1		
ID	Q9ERD1	PRELIMINARY;
AC	Q9ERD1;	PRT; 218 AA.

 $\alpha$





```
RT products involved in the assembly of functional protein complexes." ;
RL Submitted (DEC-1998) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ131892; CAB40350.1; -.
DR InterPro: IPR000901; CPSase.
DR InterPro: IPR002819; HD.
DR PROSITE; PS00867; CPSASE_2; UNKNOWN_1.
SQ SEQUENCE 3616 AA; 418996 MW; 6D3BB330B6BCA92A CRC64;

Query Match 6.4%; Score 161; DB 13; Length 3616;
Best Local Similarity 19.5%; Pred. No. 0.033;
Matches 105; Conservative 88; Mismatches 209; Indels 136; Gaps 22;

Qy 4 WDYSNGMKPLPVLVIFVCLLWLKDSHCA-----PTWK---DKTAISENLKSF 48
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1479 WDTEAEDSVKFSIHVG-----YKES-CSSVHGSSMGDDDDITMWSGTAEEGLLSQHL--- 1529

Qy 49 SEVG-EIDADEE-----VKALGIGIKQKIMMERKEKEHTNLMSTLKKREE 94
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1530 AESGVLPDENEELVNLSSRLQAAVEKLLLEAINETSNOLEHAKITQETLMRESFKKQEE 1589

Qy 95 KOEALKLLNEVQEHLEERLCRESLA-----DSWGECSRLENMCMR---IYTT 141
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1590 ATEFIRYQEEQLERLSEETK-AREQLALELNKAEGLDGYADEKAFLEKQLQEKIDVIDH 1648

Qy 142 COPSWSYVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLI--EKDAQLTQMEDVFSOLT 199
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1649 LEQELLCGKNLQEL-----BAEQQQIQEERKELLARQKADAMRADAPVEQQLL 1696

Qy 200 VDVNSLFNRSNVFMQOEEDTQSFHSFDITDLETPY--FFPAFSKEPMTKADLEQCW 257
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1697 ETEKLMKEIEVQQAKEYEDDDQKQVKVLEIDLEEQVSRFIELEQEKNAELMDLRO-- 1754

Qy 258 DIPNFFQLFCNFVSIVYSVSETITKML-----KAIEDLPKQDKAPDHGG 302
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1755 -----QNQALEKQLEKTRKFLDEQAVDREHERDVFEQETQLEQOLKVPORSQ 1802

Qy 303 LISKMLPGDRLGCELDQNLRSRCKFEKQKCOAHLSEDCPDVPALH---TELDEAIR 359
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1803 PVNEHQSRVEQOLTWHLKEKTDKCSLELLSRKFLQORDIQERNEETKELECRIRLEQAL- 1861

Qy 360 LVNVSN-----QYGOIL-----QMTKRHLEDATVLMKRGQFGCW 396
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1862 IISANDLQVKERKQFGIIVKGLPFLQIQQAEREAVDRKEKVTN--LEEQLEQFR-- 1917

Qy 397 SELANQAPETEIFNSIQV-----VPRHEGNISKQDETMMYDLSILPSSNFTLK 446
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 1918 BELENKNEEVQOLHMQLEIQRKESTTHLQLEQENKLFKDEMEILGLAIQKSEDTIK 1975

RESULT 11
O66878 PRELIMINARY; PRT; 1156 AA.
AC O66878;
DT 01-AUG-1998 (TremBLrel. 07, Created)
DT 01-AUG-1998 (TremBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE CHROMOSOME ASSEMBLY PROTEIN HOMOLOG.
GN XCPC OR AQ.629.
OS Aquifex aeolicus.
OC Bacteria; Aquificales; Aquificaceae; Aquifex.
OX NCBI_TaxID=63363;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=VF5;
RX MEDLINE=98196666; PubMed=9537320;
RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Auway M., Huber R.,
RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT "The complete genome of the hyperthermophilic bacterium Aquifex
RL aeolicus." ;
RL Nature 392:353-358(1998).
DR EMBL; AE000699; AAC06839.1; -.

DR InterPro: IPR003439; ABC_transportr.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003716; RNA_pol_omega.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR InterPro: IPR002017; Spectrin.
DR Pfam; PF02483; SMC_C; 1.
DR Pfam; PF02463; SMC_N; 1.
DR PROSITE; PS00211; ABC_TRANSPORTER; UNKNOWN_1.
KW Complete proteome.
SQ SEQUENCE 1156 AA; 135563 MW; B12DB30F70COCE49 CRC64;

Query Match 5.9%; Score 150; DB 16; Length 1156;
Best Local Similarity 20.3%; Pred. No. 0.049;
Matches 90; Conservative 88; Mismatches 162; Indels 104; Gaps 17;

Qy 45 LKSESEGEID-----ADEEVKALTGKQMKIMWE-----RKEKEHTNLMSTLKK 90
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 163 IETSGIGEYERKKEKALELAELVELKIKELDILEISNQLKRLKEKEKLEKFKELQR 222

Qy 91 CREEQBALKLLNEVQEHLEERLCRE--SLADSWGECSRLENMCMRYTTTCQPSWS 148
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 223 IKRE-TEAKILLKEKELRERILNELSLRESLEDITFQIQENKEL----- 271

Qy 149 VKNIERFRKIYQFLPFPHEDNEK---DLPISSEKLI-EKDAQLTQMEDVFSOLTVDVNS 204
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 272 --NERERILLKEVNEKIMPFEKVGKFTAEIENAESIKERELKESENRYKNLEELINN 329

Qy 205 LFNSRNVFRO--MQEEDTQSFHSFDITDLETPY--FFPAFSKEPMTKADLEQCW 260
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 330 LLSKENLERVGTQLELEK-----LKEEYKSLKEVEREKLRELEEE----- 373

Qy 261 NFFQLFCNF-SVSIYVSSETITKMLKAIEDLPKQDKAPDHGGGLISKMLPGQDRGLCGEL 319
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 374 --ERLKITFDVVKLEEKELTEKLSL-NKEQLELIQANLNKI-----ERI 421

Qy 320 DQNLRSRCKFEKQKCOAHLSEDCPDVPALHTELDEAIRLVNVSNQYGOILQMTKRHL 379
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 422 KEDINKLISREEEKIKETKEQELKRLKAIKKEEBELRNLTOELNIYERLSEVRKKL 481

Qy 380 E-----DTAYLVKMKRGQFGVSELAN-QAPE----- 405
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 482 EEVLKKEGAIEREVRSDVDVFKDKGVGVSELRVKNPEHITAIEVAGGRLKFI 541

Qy 406 ----TEIIFNSIQVPRHEGNIS 425
|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
Db 542 VVEDEEVAKECIQLAKRMNLGRFS 565

RESULT 12
Q95JRO PRELIMINARY; PRT; 840 AA.
AC Q95JRO;
DT 01-DEC-2001 (TremBLrel. 19, Created)
DT 01-DEC-2001 (TremBLrel. 19, Last sequence update)
DE 01-DEC-2001 (TremBLrel. 19, Last annotation update)
DE HYPOTHETICAL 97.6 KDA PROTEIN.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopitheciidae;
OC Cercopitheciinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=TESTIS;
RA Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai M.,
RA Terao K., Sugano S.;
RT "Isolation of novel full-length cDNA clones from macaque testis cDNA
RL libraries." ;
RL Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AB070119; BABG3064.1; -.
KW Hypothetical protein.
```

SQ SEQUENCE 840 AA; 97587 MW; A875627BE3D4D716 CRC64;

Query Match 5.9%; Score 149.5; DB 6; Length 840;

Best Local Similarity 18.6%; Pred. No. 0.036; Mismatches 154; Indels 143; Gaps 18;

Matches 88; Conservative 89; Mismatches 154; Indels 143; Gaps 18;

QY 39 TAISENLKSFSEVGEIDADEEYKALGTGKIMMERKEKHTNLMSTLKKCRE-EKQE 97

DB 426 TEMQKNKSVQY--LEMDKTLKKEEVEKRLQQLRKQEKVYASALDLLKREKETQOE 483

QY 98 ALKLLNEVQE---HLEERELCRSLADSWGECRSCLENNCMRIYTCQPSWSSVKNKI 153

DB 484 FLSLQEEFQKRDKANLEERQKL-----KSRLL 509

QY 154 ERFFRKIYQFLPFHEDNEKDLPISEKLEKDAQLTOMEDVFSQLTVDVNSLNRSE--- 210

DB 510 EKLTLQVKNLQPMSENERAKNKLQO-----QINEVKNENKLLQHVARSEEQNYVPK 562

QY 211 NVFRMQQEFQDTFQSHFISDTLTPYFPFAPS---REPMTKADLE----- 254

DB 563 SETAQLKEQLEEVKMSDITKTKMTHSNLLDSCPEEESLAPADIERSQLASKMHSLL 622

QY 255 -----QCWDIPNPFQFCNFSVSIYESVETITKMLKAI-----EDLPKQDKAP 298

DB 623 ALMVGILLKQODITNSDAE--HFKESS--EKVSDIMLQRLKSLHKKKLLDKKLLKHDRIT 678

QY 299 DHGGLSKMLPGDRLG-----CGE----- 318

DB 679 TFRDLIAKEKAFQDRAIKVTDSCDSDEAKSIRDVPTFLGAKLKYHSLNEELDFLITKLGC 738

QY 319 -LDQNLRSRCFKFEKQKQAHLSDECDVPALHTDELDAIRLVNVSNOYQOILQMTKR 377

DB 739 LLESKESHCRNLEENDKYORHLGSLIKKVTSEYEEIECADQRLAISHSQIAH-LEKRNK 797

QY 378 HLEDYALIVEKMRGQGWVSELANQAPETEIFN---STQVPRHNEGHSKQD 428

DB 798 HLED---LIRKPR-----EKARK-PRSKSLENHPKSMTMPAVFKNRNDLD 840

RESULT 13

Q28021

ID Q28021 PRELIMINARY; PRT; 1388 AA.

AC Q28021;

DT 01-NOV-1996 (TremBLrel. 01, Created)

DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)

DE RHO-ASSOCIATED KINASE.

OS Bos taurus (Bovine).

OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;

OC Bovidae; Bovinae; Bos.

OX NCBI\_TaxID=9913;

RN [1]

RP SEQUENCE FROM N.A.

RX MEDLINE=96208507; PubMed=8641286;

RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,

RA Nakano T., Okawa K., Iwamatsu A., Kaibuchi K.;

RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative

RL target for small GTP binding protein Rho.";

CC EMBO J. 15:2208-2216(1996).

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.

DR EMBL; U36909; AAC48567.1; -

DR HSSP; Q63450; 1A06.

DR InterPro; IPR002219; DAG\_PE-bind.

DR InterPro; IPR000719; Euk\_pkinase.

DR InterPro; IPR001849; PH.

DR InterPro; IPR000961; Pkinase\_C.

DR InterPro; IPR000861; REM\_repeat.

DR InterPro; IPR002290; Ser\_thr\_pkinase.

DR Pfam; PF02185; DAG\_PE\_bind; 1.

DR Pfam; PF02185; HRI; 1.

DR Pfam; PF00169; PH; 1.

pfam; PF00069; pkinase; 1.  
DR SMART; SM00109; C1; 1.  
DR SMART; SM00074; HRI; 1.  
DR SMART; SM00233; PH; 1.  
DR SMART; SM00220; S-TKC; 1.  
DR SMART; SM00133; S-TK\_X; 1.  
DR PROSITE; PS00081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE; PS00003; PH\_DOMAIN; 1.  
DR PROSITE; PS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE; PS00111; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE; PS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 1388 AA; 160799 MW; CA83CE7D3860465D CRC64;

Query Match 5.9%; Score 149; DB 6; Length 1388;

Best Local Similarity 16.7%; Pred. No. 0.071;

Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;

QY 36 KDKTAISENLKSFSEVGEIDADEEYKALGTGKIMKI-----MMERKEKEHTNLM 86

DB 760 KOKDVLNEDVRNLT---LKIEQETQKRLTQNDLKMOTQOVNTLMSEKQLKQENHLL 835

QY 87 TLK-----KCREKQKALKLNEVQEHLEEE-----RLCR 117

DB 836 ENKMSLEKQNAELRRQDADQGMKQLOLEAEQYFSTLYKTQVRELKEECEKTKLCK 895

QY 118 E-----SLADSWGECRSCLENNCMRIYTCQPSWSSVKNKIERFFRKIYQFLFPFHEDN 171

DB 896 ELQKKQKQELQDE-----RDSLAQLEITLTKADSEQLARSTAEQYSDLEKE 942

QY 172 E--KDLPISE-----KLIEXDAQLTQMEDVFSQLTVDVNSLFRNPNVFRMQOEFQ 222

DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLTSVANLANEKEELNNKLKQAO 1002

QY 223 TQSHFISDTLTPYFPFAPSKEPMTKADLEQCWDIPNFFQFCNFSVSIYESVETIT 282

DB 1003 LSR---LKDEEISAAAIKAQFEKOLLTERLK-----TQAVN 1036

QY 283 KMLKAIE-----DLPKQDKAPDHGGLSKMLPGDRLGQEL---DQNLRSRCF 327

DB 1037 KLAETMRKEPVKRGNDTQVRKEK-----ENRKLHMLKSKEREKLTQMI 1082

QY 328 KFHEKQKQAHLSDECDVPALHTELDEAIRLVNVSNOYQOILQMTKRHLE----- 380

DB 1083 KYOKELNEMQAIAEE---SOIRTELQMTLDSKDSIEQLRSQALHIGLDSSSISGS 1138

QY 381 -----DTAYLIVEKMRG-----QFGWVSEL----- 399

DB 1139 PGDTEADGFPESRLEGWLSLVRNNTKFGWVKYIVVSSKKILFYDSEQDKESNPYM 1198

QY 400 -----ANQAPETEIFNSIQVVPRI-----HEGNISKQDEMTMTDLSTLPSNFT 444

DB 1199 VLDIDKLFHVRPVTQDQVYRADAKEIPRFQILYANEGESKKEQE----- 1243

QY 445 LKIPLEESAESSNFI---GYVVAKALQHF 470

DB 1244 --FPVEPVGEKSNYICHKGHEFIPTLYHF 1270

RESULT 14

Q18082

ID Q18082 PRELIMINARY; PRT; 1091 AA.

AC Q18082;

DT 01-NOV-1996 (TremBLrel. 01, Created)

DT 01-NOV-1996 (TremBLrel. 01, Last sequence update)

DT 01-DEC-2001 (TremBLrel. 19, Last annotation update)

DE HYPOTHETICAL 127.0 KDA PROTEIN.

GN C18C4.5.

OS Caenorhabditis elegans.

OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;

OC Rhabditidae; Peloderinae; Caenorhabditis.

OX NCBI\_TaxID=6239;

RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX MEDLINE=99069613; PubMed=9851916;  
RA None;  
RT "Genome sequence of the nematode C. elegans: a platform for  
investigating biology. The C. elegans Sequencing Consortium. ";  
RL Science 282:2012-2018(1998).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Gattung S.;  
RA "The sequence of C. elegans cosmid C18C4. ";  
RT Submitted (APR-1996) to the EMBL/GenBank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RC STRAIN-BRISTOL N2;  
RX Waterston R.;  
RA "Direct Submission. ";  
RT Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.  
RL EMBL; U55369; AAK52177.1; -;  
DR HSSP; P42227; 18G1.  
KW Hypothetical protein.  
SQ SEQUENCE 1091 AA; 127046 MW; C3EC4AE14E962132 CRC64;

Query Match 5.9%; Score 148.5; DB 5; Length 1091;  
Best Local Similarity 20.3%; Pred. No. 0.057;  
Matches 104; Conservative 100; Mismatches 184; Indels 125; Gaps 25;  
  
QY 36 KDKTAISNLKSFSEVEID-----ADEVKKALTGKQMK--IMMERKEKEHTNLM 85  
DB 283 KDNAIL-NVQLRDKDKIDRIQVDLAAESRAQQAEDVRDKERITTSKKDDSDNNLL 341  
QY 86 -STLKRCRQEAALKLNEVQEHLEEEERLCRESLADSWGECRSLNNCMR-----IV 139  
DB 342 QDELRRTEEKYQQAQKTIENLDEIKQEQTIQIRD-LGRSLDEAKRQLQKSEQRQNEVA 400  
QY 140 TTCOPSSVKNKTERFRKIYQFLFPFHEDNEKDLPISEKLIKDAQLTQ----- 190  
DB 401 RQGEDSARSMEEKATKEEIKLKKSQVQLQQLQDLQKRVQ--ELTEQRKVLESKA 457  
QY 191 -MEDVFSQITVDVNSLFRNFVFRMQQEF-----DQTEQ----- 225  
DB 458 SVADEFGTLASSLNSL--REEN--RQVEETRSLOTNIRILODEVYQHQDAITEWKNRAE 513  
QY 226 -----SHFIS-----DITDTEPYFPFAFESKPMTKADLEQCWDIPNFFQLFCNFS 270  
DB 514 KAEYIEKENHRVQVNSSHSDADITRLENEKTOEEALEKADQEQQAIRE-----ASES 568  
QY 271 VSIYE---SVSETITKMLKAIEDLPKQKAPDHGGGLSKMLPGQDRGLCGELDONLSRC 326  
DB 569 VRVKNREMTASITSDRIQSLKE--KVDS-----LTRELESSRRM-EQLQEDQTKF 618  
QY 327 FKPEKQKQQAHLSEDCPVPALHTELDEAIRLVNSNQYQGIQLQMTKRKHELDYALV 386  
DB 619 LGSND---ETKAEMMKD-----LHEAQDETEKLTNQAQGLSKNETLTTE-LDSQNL 668  
QY 387 EKMKGQGWV---SELANAQAPETEIFNFIQVPRTHGNIKSKQDPTMTDLSILPSSN 442  
DB 669 ERKQAQYKADKYEETKVLQREAEDLADRLQAA-QILSGNVESKFSMDQKE----- 719  
QY 443 FTLKIPLEESAESSNFIQYVYVAKALQHFKEHK 475  
DB 720 --SKIEMERILDNNH-----KELEKLREELK 743  
  
RESULT 15  
Q14789  
ID Q14789 PRELIMINARY; PRT; 3259 AA.  
AC Q14789; Q14398;  
DT 01-NOV-1996 (TrEMBLrel. 01, Created)  
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)

DT 01-MAY-2000 (TrEMBLrel. 13, Last annotation update)  
DE GIANTIN (GCP372) (MACROGOLGIN) (GOLGI AUTOANTIGEN, GOLGIN SUBFAMILY B,  
DE 1).  
GN GOLGB1.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94187728; PubMed=7511208;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Molecular genetic analyses of a 376-kilodalton Golgi complex membrane  
protein (giantin) [retracted in Mol Cell Biol 1995 Jan;15(1):591].";  
RL Mol. Cell. Biol. 14:2564-2576(1994).  
RN [2]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94257116; PubMed=8198703;  
RA Seelig H.P., Schranz P., Schroeter H., Wiemann C., Griffiths G.,  
RA Renz M.;  
RT "Macrogolgin--a new 376 kD Golgi complex outer membrane protein as  
target of antibodies in patients with rheumatic diseases and HIV  
infections. ";  
RL J. Autoimmun. 7:67-91(1994).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=95100974; PubMed=7802676;  
RA Sobda M., Misumi Y., Fujiwara T., Nishiohara M., Ikehara Y.;  
RT "Molecular cloning and sequence analysis of a human 372-kDa protein  
localized in the Golgi complex. ";  
RL Biochem. Biophys. Res. Commun. 205:1399-1408(1994).  
CC -1- FUNCTION: MAY PARTICIPATE IN FORMING INTERCISTERNAL CROSS-BRIDGES  
OF THE GOLGI COMPLEX.  
CC -1- SUBUNIT: DISULFIDE-LINKED HOMODIMER.  
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. GOLGI MEMBRANE.  
CC -1- ALTERNATIVE PRODUCTS: DIFFERENTS ISOFORMS MAY ARISE BY ALTERNATIVE  
SPLICING.  
CC -1- DISEASE: ANTIGEN IN CHRONIC RHEUMATOID ARTHRITIS AND IN THE  
AUTOIMMUNE DISEASE SJOEGREN'S SYNDROME.  
DR EMBL; X75304; CAAS3052.1; -;  
DR EMBL; D25542; BAA05025.1; -;  
DR MIM; 602500; -;  
KW Golgi stack; Antigen; Coiled coil; Transmembrane;  
KW Alternative splicing.  
FT DOMAIN 1 3235 CYTOPLASMIC (POTENTIAL).  
FT TRANSMEM 3236 3256 POTENTIAL.  
FT DOMAIN 3257 3259 LUMENAL (POTENTIAL).  
FT DOMAIN 48 110 COILED COIL (POTENTIAL).  
FT DOMAIN 127 223 COILED COIL (POTENTIAL).  
FT DOMAIN 238 448 COILED COIL (POTENTIAL).  
FT DOMAIN 460 526 COILED COIL (POTENTIAL).  
FT DOMAIN 545 593 COILED COIL (POTENTIAL).  
FT DOMAIN 677 956 COILED COIL (POTENTIAL).  
FT DOMAIN 969 1028 COILED COIL (POTENTIAL).  
FT DOMAIN 1062 1128 COILED COIL (POTENTIAL).  
FT DOMAIN 1154 1245 COILED COIL (POTENTIAL).  
FT DOMAIN 1301 1779 COILED COIL (POTENTIAL).  
FT DOMAIN 1828 2781 COILED COIL (POTENTIAL).  
FT DOMAIN 2797 2857 COILED COIL (POTENTIAL).  
FT DOMAIN 2872 2993 COILED COIL (POTENTIAL).  
FT DOMAIN 3026 3102 COILED COIL (POTENTIAL).  
FT DOMAIN 3133 3185 COILED COIL (POTENTIAL).  
FT DOMAIN 2420 2423 COILED COIL (POTENTIAL).  
FT DOMAIN 2993 POLY-SER.  
FT VARSPLIC 1 39 MISSING (IN REF. 3).  
FT CONFLICT 215 215 A -> AQLSSM (IN REF. 3).  
FT CONFLICT 1765 1765 D -> G (IN REF. 3).  
FT CONFLICT 2950 2950 H -> D (IN REF. 3).  
SQ SEQUENCE 3259 AA; 376075 MW; 60376A20D8A178DD CRC64;

Query Match

5.8%; Score 147; DB 4; Length 3259;

Best Local Similarity 19.5%; Pred. No. 0.27;  
Matches 104; Conservative 96; Mismatches 167; Indels 166; Gaps 23;

QY 44 NLKSFSEGEIDADEE-VKALTGKQMKIMMERKEKHTNLMSTLKK-----CREEKQEA 98  
Db 2378 NMKEQKIIISLGSKEEAIQVAIAELRQOH---DKEIKELNLSQEEENIVLEENKKA 2434  
QY 99 LKLLNEVQEHLE--EEERLCRESLADSGECRSCLENNCMRI---YTTQCPSSSV---- 149  
Db 2435 VDKTNOLMETLTKIKENTQQAQDLSFVKSSNDNRDRIVGDYQQLERHLSIILEK 2494  
QY 150 -----KNKIERFFRKYYQFLPPFHEDNEKDLPISEKLE-----KDAQL 188  
Db 2495 DOLIOBAAAENKKEEIRGLRSHMDLNSNAK---LDAELIQYREDLNOVITIKDSQQ 2551  
QY 189 TQ-----MEDVFSQL-----TVDVNSLNRSENVFQMQQEPDQTFQSHFISD 231  
Db 2552 KOLLEVOLOONKELENKYAKLEEKLEKESEANEDLRRSFNALQEEKQDLSKEIESLKVSI 2611  
QY 232 TDLT-----EPYFFPA-FSKEPMTKADLEOCWDIPNFFQLF 266  
Db 2612 SOLTROVTALQEBGLGLYHAQLVKKEEVHRLSALFSSQKRIAELEE-----ELV 2663  
QY 267 CNFSVSIYESVSETITPKMLKAIED-LPKQDKAPDHGGLISK----- 306  
Db 2664 C-----VQEAARKVGEIETDKLKKELKHLHDAGIMRNETETAERVAELARDLVE 2714  
QY 307 -----MLPGQDRGLCGELDONLSRCFKPEKQCKQCOAHLSEDCPDVPALHTELDEAIRL 360  
Db 2715 MEQKLLMVTKENKGLTAQI-QSFGISM-----SSLQNSRDHANE-----ELDELKRRK 2760  
QY 361 VNVSNQOYGQIILQMTKRKHLEDYAYLVKMRGOGFWGVSSELANQAPETEIIFNSTQVVPRTH 420  
Db 2761 YDASLKELAQLEQGLNRRDALL-----SETAFSMNSTE--ENSLSHLEKLN 2807  
QY 421 EGNISKODETMMTDLSILPSSNFTLKIPLEESAESSNFTGYVYVAKALQHKEH 473  
Db 2808 QOLLSKDEQLL-----HLSSQLEDSYNQVQSFSKAMA-SLQNERDH 2847

Search completed: July 2, 2002, 11:57:14  
Job time: 412 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:32:26 ; Search time 56.55 seconds  
(without alignments)  
972.264 Million cell updates/sec

Title: US-09-722-544A-2  
Perfect score: 2632  
Sequence: 1 MKIKAEKNEGSPRSWWQLHW.....FIGYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 11073796 residues  
Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*  
1: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDSI/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2586	98.3	495	AA130785	Protein encoded by
2	2464	93.6	466	AA130793	Immature human HKN
3	2447	93.0	477	AA130786	Protein encoded by
4	2350	89.3	446	AA130792	Mature secreted hum
5	1757.5	66.8	465	AA130791	Bovine HKNG1 ortho
6	1639	62.3	466	AA130787	Protein encoded by
7	1558	59.2	450	AA130788	Guinea pig HKNG1 o
8	1277	48.5	374	AA130789	Guinea pig HKNG1 o
9	1272.5	48.3	373	AA130790	Guinea pig HKNG1 o
10	925.5	35.2	521	AA130794	Amino acid sequenc
11	434	16.5	449	AA128048	Novel human secret

12	432	16.4	448	12	AA11704	Cytolysis Inhibito
13	405.5	15.4	416	22	ABBS0285	Apolipoprotein J o
14	400.5	15.2	416	21	AA803441	Porcine clusterin
15	320	12.2	363	22	AAE03764	Human gene 1 encod
16	173.5	6.6	148	22	AAO00607	Human polypeptide
17	160.5	6.1	116	21	AAO30745	Human secreted pro
18	159.5	6.1	139	22	AAE03783	Human gene 1 encod
19	159.5	6.1	139	22	AAE01703	Human gene 4 encod
20	152.5	5.8	247	22	AAE03787	Human gene 1 encod
21	145	5.5	1372	19	AAW56473	Protein with Rho p
22	143	5.4	1931	22	ABE61012	Drosophila melanog
23	142.5	5.4	2633	22	ABG06505	Novel human diagno
24	141.5	5.4	3433	18	AAW22017	utrophin. Homo sa
25	141	5.4	944	21	AAV67600	Human adipose tiss
26	141	5.4	1388	19	AAW56475	Protein with Rho p
27	139.5	5.3	2663	22	AAW39097	Human polypeptide
28	139.5	5.3	2688	22	AAW40883	Human polypeptide
29	139	5.3	1374	22	AAW69070	Human male enhance
30	136	5.2	934	22	AAU01768	Human secreted pro
31	136	5.2	2517	21	AAV71159	Human phosphodiast
32	135.5	5.1	1427	12	AA10534	Human 160kb mediat
33	134	5.1	5373	22	AAU14603	Novel bone marrow
34	134	5.1	5447	22	AAU14697	Human polypeptide
35	130.5	5.0	1788	22	AAW40467	Human mitosis. Ho
36	130.5	5.0	2482	16	AAW72826	Human mitosis. Ho
37	130.5	5.0	2482	19	AAW23996	Human mitosis amin
38	130	4.9	990	22	AAW78520	Human protein SEQ
39	129	4.9	816	16	AAW66931	AMML chromosome in
40	128.5	4.9	1392	20	AAW06999	Restin protein seq
41	128.5	4.9	1411	17	AAW02258	Nucleolar/endosoma
42	128.5	4.9	3248	17	AAW99795	Kinetochore protei
43	128	4.9	885	16	AAW66930	AMML chromosome in
44	127	4.8	1047	18	AAW01535	Cellular homologue
45	126.5	4.8	1780	22	AAW38681	Human polypeptide

ALIGNMENTS

RESULT 1  
AA130785  
ID AAY30785 standard; Protein: 495 AA.  
XX  
AC AAY30785;  
XX  
DT 23-NOV-1999 (first entry)  
XX  
DE Protein encoded by human HKNG1 cDNA.  
DE  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
PF 16-MAR-1999; 99WO-US05606.  
XX  
PR 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX  
(MILL-) MILLENNIUM PHARM INC.  
(REGC) UNIV CALIFORNIA.  
PA  
PI Chen H, Freimer NB;  
XX  
DR WPI; 1999-562047/47.  
DR N-PSDB; AA210750.





QY 150 ENNCMIYTTCPSSVSKNIERFRKIYQFLFPFHEDNEKDLPISEKLEEDAQLTQM 209  
 Db 121 enncmriyttcpsvsvknierfrkiyqflfpfhednekdipisekleeaqltqm 180  
 QY 210 EDVFSQLVDSVNSLFRNSFNVRQMQQEFDTQFSHFISDTDLTEPYFFPAFSPKMTKA 269  
 Db 181 edvfsqltdvnslnfrsnfrvmqgfdqtfqshfisdtdltepyffafskpmtka 240  
 QY 270 DLEQCWDIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHPGGLISKMLPGQ 329  
 Db 241 dleqcwdipnffqfcnfsvsiyesvettikmlkaiedlpkqdkapdhpghglskmlpgq 300  
 QY 330 DRGLGELDQNLNRFCFHEKQKQCAHLSEDCPDVPAHTELDIAIRLVNVSQYQGI 389  
 Db 301 drglcgeldqlnsrckfhekckcqhalsedcpdvpalhtelddeairlvnvsqyqgi 360  
 QY 390 LQMRKHELDYALVLEKMRGQFGWVSELANQAPETIIFNSIQVVPRIHEGNISKQDET 449  
 Db 361 lqmrkheldtayiveknrgqfgwvselanqapetiiifnsiqvvprihegniskqdet 420  
 QY 450 MTDLSILPSSNFTKIPLEESAESSNFYGYVVAQALQHFKEHFKTW 495  
 Db 421 mtdlsilpssnftkipleesaessnfygyvvakalqhfkehfktx 466

RESULT 3  
 AAY30786  
 ID AAY30786 standard; Protein: 477 AA.  
 XX  
 AC AAY30786;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Protein encoded by a human HKNG1 splice variant HKNG1-V1.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-US05606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 PR 05-JUN-1998; 98US-0088312.  
 PR 28-OCT-1998; 98US-0106056.  
 PR 22-JAN-1999; 99US-0236134.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chen H, Freimer NB;  
 XX  
 DR WPI; 1999-562047/47.  
 DR N-PSDB; AAZ10751.  
 XX  
 PT New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia -  
 XX  
 PS Claim 1; Fig 2A-B; 205pp; English.  
 XX  
 CC The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)  
 CC splice variant. HKNG1 is a gene associated with bipolar affective  
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
 CC activity in individuals can then be administered therapeutically to  
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders

CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 477 AA;

Query Match 93.0%; Score 2447; DB 20; Length 477;  
 Best Local Similarity 99.1%; Pred. No. 4.9e-194;  
 Matches 466; Conservative 1; Mismatches 3; Indels 0; Gaps 0;  
 QY 25 NNSGNMKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKKALTG 84  
 Db 7 sngnmkpplllvfivclllwlkdschaptwkdktaisenlksfsevgeidadeevkkaltg 66  
 QY 85 IKOMKIMMERKEKEHTNLMSTLKKCREKQKALKLLNEVQEHLEERLCRESLADSMGE 144  
 Db 67 ikqmkimmerkekehtnlmstlkkcreekqaalkllnevqehleeeerlcresladswe 126  
 QY 145 CRSCLENNCMIYTTCPSSVSKNIERFRKIYQFLFPFHEDNEKDLPISEKLEIEDA 204  
 Db 127 crsclenncmriyttcpsvsvknierfrkiyqflfpfhednekdipisekleieeda 186  
 QY 205 QLTOMEDVFSQLVDSVNSLFRNSFNVRQMQQEFDTQFSHFISDTDLTEPYFFPAFSKE 264  
 Db 187 qltqmedvfsqltdvnslnfrsnfrvmqgfdqtfqshfisdtdltepyffpfske 246  
 QY 265 PMTKADLEQCWDIPNFFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHPGGLISK 324  
 Db 247 pmtkadleqcwdipnffqfcnfsvsiyesvettikmlkaiedlpkqdkapdhpghglsk 306  
 QY 325 MLPQDRLGCLGELDQNLNRFCFHEKQKQCAHLSEDCPDVPAHTELDIAIRLVNVSQ 384  
 Db 307 mlpgdrglclgeldqlnsrckfhekckcqhalsedcpdvpalhtelddeairlvnvsq 366  
 QY 385 QYGQILQMRKHELDYALVLEKMRGQFGWVSELANQAPETIIFNSIQVVPRIHEGNISK 444  
 Db 367 qygqilqmrkheldtayiveknrgqfgwvselanqapetiiifnsiqvvprihegnisk 426  
 QY 445 QDETMMTDLISLPSSNFTKIPLEESAESSNFYGYVVAQALQHFKEHFKT 494  
 Db 427 qdetmmtdlsilpssnftkipleesaessnfygyvvakalqhfkehfktx 476

RESULT 4  
 AAY30792  
 ID AAY30792 standard; Protein: 446 AA.  
 XX  
 AC AAY30792;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Mature secreted human HKNG1 protein sequence.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN W09947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX

PF	16-MAR-1999;	99WO-US05606.	
XX			
PR	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC ) UNIV CALIFORNIA.		
XX			
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia -		
XX			
PS	Claim 5; Fig 17; 205pp; English.		
XX			
CC	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1		
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	polynucleotides are useful to identify compounds modulating HKNG1 gene		
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting		
CC	or enhancing HKNG1 gene expression or activity in individuals can then		
CC	be administered therapeutically to treat HKNG1-mediated disorders,		
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or		
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal		
CC	dominant myopia. The polynucleotides can be used in gene therapy		
CC	techniques to treat such disorders. They are also useful in diagnosis		
CC	to identify individuals having, or at risk of developing, HKNG1-mediated		
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially		
CC	result in the production of a protein with a different sequence to		
CC	the human full-length HKNG1 polypeptide or splice variant sequence,		
CC	especially the substitution of a lysine for a glutamic acid at residue		
CC	202 or 184. The polynucleotides are also useful in gene mapping, to		
CC	produce probes or primers to identify similar sequences (e.g. mutants		
CC	or sequences from different species) and to produce transgenic		
CC	animals.		
XX			
SQ	Sequence 446 AA;		
<hr/>			
Query Match 89.3%; Score 2350; DB 20; Length 446;			
Best Local Similarity 100.0%; Pred. No. 4.8e-186;			
Matches 446; Conservative 0; Mismatches 0; Indels 0; Gaps 0;			
QY	50	APTWKDKTAISENLKSFSEYGEIDADEVKKALTGKQKIMMERKEKHTNLMSTLKKC	109
Db	1	aptwkdtalsenlkfseveygeidadeevkaltgikqkimmerkekehtnlmstlkkc	60
QY	110	REEKQKALKLLNEVQEHLEERLCRESLADSWGECRSCLENMCIYTTTCQPSWSVKN	169
Db	61	reekqealkllnevqehleerlcresladswegecrscleennmciytttcqpswsvkn	120
QY	170	KIEREFKIKYQFFPHFHEKDLPISEKLEEDAQLTQMEDVFSQLTVDVNSLFRSFN	229
Db	121	kierfrkiyqffphfhekdplisekleedaqltqmedvqsltdvsnlfrsfn	180
QY	230	VFRMQQEFQDTQSFHSISDITETPEYFFPAFSKEPMTKADLEQCWDIPNFFQLFCNFSV	289
Db	181	vfrmqgefqtqsfhsisdltetpyffpafskpmtkadleqcwdipnffqlfcnfsv	240
QY	290	SIYESVSETTKMLKAIEDLPKQKAPDHGGLISKMLPGODRGLCGELDONLSRCFKFHE	349
Db	241	siyevsettkmlkaiedlpkqkaphdggliskmlpgqdrglcgeldnlsrckfhe	300
QY	350	KCKQCAHLSEDPCDVPALHTEDEATRLVNSNQYQGTLOMTRKHLDETAYLVEKMRG	409
Db	301	kckqcahlsedcpdvpalhhteidearlvnvnqyqgltomtrkhldeatylvekmrg	360
QY	410	QFGWSELANQAPETEIIFNISQVVPRIHEGNISKQDETMMTDLISILPSSNFTLKIPLEE	469

Db	361	qfgwselanqapeteiifnsiqvvprihegniskqdetmmtdlslpssnftlkiplee	420
QY	470	SAESSNFIGYVVAQALQHKEHFKTW	495
Db	421	saessnfigyvvaqalqhkehfktw	446
<hr/>			
RESULT	5		
AY30791			
ID	AA30791 standard; Protein; 465 AA.		
XX			
AC	AA30791;		
XX			
DT	23-NOV-1999 (first entry)		
XX			
DE	Bovine HKNG1 ortholog splice variant protein.		
XX			
KW	HKNG1: Hong Kong new gene 1; bipolar affective disorder; BAD;		
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;		
KW	schizophrenia; splice variant.		
XX			
OS	Bos sp.		
XX			
PN	WO9947535-A1.		
XX			
PD	23-SEP-1999.		
XX			
PF	16-MAR-1999; 99WO-US05606.		
XX			
PR	16-MAR-1998; 98US-0078044.		
PR	05-JUN-1998; 98US-0088312.		
PR	28-OCT-1998; 98US-0106056.		
PR	22-JAN-1999; 99US-0236134.		
XX			
PA	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC ) UNIV CALIFORNIA.		
XX			
PI	Chen H, Freimer NB;		
XX			
DR	WPI; 1999-562047/47.		
DR	N-PSDB; AA210759, AA210760, AA210761.		
XX			
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia -		
XX			
PS	Claim 1; Fig 11A-B; 205pp; English.		
XX			
CC	The present sequence is encoded by bovine HKNG1 ortholog splice		
CC	variant. HKNG1 (Hong Kong new gene 1) is a gene associated		
CC	with bipolar affective disorder (BAD). HKNG1 polynucleotides are		
CC	useful to identify compounds modulating HKNG1 gene expression or		
CC	HKNG1 polypeptide expression/activity. Compounds inhibiting or		
CC	enhancing HKNG1 gene expression or activity in individuals can		
CC	then be administered therapeutically to treat HKNG1-mediated		
CC	disorders, especially neuropsychiatric disorders e.g. BAD,		
CC	schizophrenia, or HKNG1-mediated myopia disorders, such as		
CC	early-onset autosomal dominant myopia. The polynucleotides can be used		
CC	in gene therapy techniques to treat such disorders. They are also useful		
CC	in diagnosis to identify individuals having, or at risk of developing,		
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such		
CC	mutations especially result in the production of a protein with a		
CC	different sequence to the human full-length HKNG1 polypeptide or		
CC	splice variant sequences, especially the substitution of a lysine for		
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also		
CC	useful in gene mapping, to produce probes or primers to identify		
CC	similar sequences (e.g. mutants or sequences from different species)		
CC	and to produce transgenic animals.		
XX			
SQ	Sequence 465 AA;		
<hr/>			
Query Match 66.8%; Score 1757.5; DB 20; Length 465;			

Best Local Similarity 71.08; Pred. No. 5.5e-137; Matches 331; Conservative 57; Mismatches 77; Indels 1; Gaps 1;	
Qy 30 MKPPLVFIIVCLLWKSHCAPTWKDKTAISENLKSFSEVGEIDAEDVEVKKALGIQOMK 89	
Db 1 mkppllvfiivylqlrdcqcaptgkdrtsiredpkgsagelvdveevkalligmqmk 60	
Qy 90 IMMERKEHTNLMSTLKKCREEQEALKLLNEVQEHLEERLCRESLADSWGECRSCL 149	
Db 61 ilmerreehsklmrtlkckreekqalklmevqehleeeerlcqyslmgswdeckscl 120	
Qy 150 ENNCMRIYTTCCPSWSSVKNKIERFRKFIYQFLPFPHEDNEKDLPISEKLEEDAAQLTQM 209	
Db 121 esdcmrfttccqsswmkstiervfrkiyqlfipfhddekelpvgekftteedvqlmqi 180	
Qy 210 EDVFSQLTVDVNSLFRNSFNVRQMQEEDOTFQSHFISDLDLTPYFFPAFSKEPMTKA 269	
Db 181 envfsqitvdvglmshfvmfkmqgcfdlafgysmtdsdmepyfifafskpaka 240	
Qy 270 DLEQCDIPNFPQLFCNFSVIYSVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329	
Db 241 hpmqswdipstfqlfcnfslsvygsaatvttemlkaiedlskqkdsahgpgssttwpvr 300	
Qy 330 DRGLCGELDQNLSCFRFHEKCKQCAHLSEDCPDVPAHTELDEATRLVNVSNQOYQOI 389	
Db 301 grglcgepgqnsseclqfharccqcdylwadcpavpelytkadealelvnlsnqgvaqv 360	
Qy 390 LQMTKRKLEDYALVERMRGQFGWVSELANQAPETEIFNSIQVVPRIHEGNSIKQDETM 449	
Db 361 lqmcpqhledtymeknreqfgwvltasqtpgsenlfsfikvpgvghgnfskqdek 420	
Qy 450 MTDLSILPSSNFTLKIPLEESAESSNFTGYGVAKALQHFKEHFTW 495	
Db 421 i-disilpssnftlitpleesaessdfisymalakavqhfkehfks 465	
RESULT 6	
AY30787	
ID AAY30787 standard; Protein: 466 AA.	
AC AAY30787;	
XX	
XX	
DT 23-NOV-1999 (first entry)	
XX	
XX	Protein enoded by the guinea pig HKNG1 ortholog gphkng1815.
DE	HKNG1: Hong Kong new gene 1; bipolar affective disorder; BAD;
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW	schizophrenia; splice variant.
XX	
OS	Cavia cobaya.
XX	
XX	W09947535-A1.
XX	
XX	23-SEP-1999.
XX	
PF 16-MAR-1999; 99WO-US05606.	
XX	
PR 16-MAR-1998; 98US-0078044.	
PR 05-JUN-1998; 98US-0088312.	
PR 28-OCT-1998; 98US-0106056.	
PR 22-JAN-1999; 99US-0236134.	
XX	
PA (MILL-) MILLENNIUM PHARM INC.	
PA (REGC ) UNIV CALIFORNIA.	
PI	
PI	Chen H, Freimer NB;
XX	
XX	WPI; 1999-562047/47.
DR	N-PSDB; AA210755.
XX	
XX	
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and

PT	schizophrenia	
XX		
PS	Claim 1; Fig 7A-B; 205pp; English.	
XX		
CC	The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.	
CC	HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective	
CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compounds	
CC	modulating HKNG1 gene expression or HKNG1 polypeptide expression/	
CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression or	
CC	activity in individuals can then be administered therapeutically to	
CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders	
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	in gene therapy techniques to treat such disorders. They are also useful	
CC	in diagnosis to identify individuals having, or at risk of developing,	
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	mutations especially result in the production of a protein with a	
CC	different sequence to the human full-length HKNG1 polypeptide or	
CC	splice variant sequences, especially the substitution of a lysine for	
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	useful in gene mapping, to produce probes or primers to identify	
CC	similar sequences (e.g. mutants or sequences from different species)	
CC	and to produce transgenic animals.	
XX		
SQ	Sequence 466 AA;	
Query Match 62.38; Score 1639; DB 20; Length 466;		
Best Local Similarity 66.78; Pred. No. 3.6e-127;		
Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;		
Qy 30 MKPPLVFIIVCLLWKSHCAPTWKDKTAISENLKSFSEVGEIDAEDVEVKKALGIQOMK 89		
Db 1 mkpllmfpvcllwkdchcceptwkdktaisenansfseaeigdvgevkialgikqmk 60		
Qy 90 IMMERKEHTNLMSTLKKCREEQEALKLLNEVQEHLEERLCRESLADSWGECRSCL 149		
Db 61 immerreehsklmktlkckeeqalklmevhehleeeslcqysladswdetrac 120		
Qy 150 ENNCMRIYTTCCPSWSSVKNKIERFRKFIYQFLPFPHEDNEKDLPISEKLEEDAAQLTQM 209		
Db 121 esncmrfttccqpsawsvknmveqfrkiyqlfplqe-ndrsgvskgvteedaqvshi 179		
Qy 210 EDVFSQLTVDVNSLFRNSFNVRQMQEEDOTFQSHFISDLDLTPYFFPAFSKEPMTKA 269		
Db 180 ehvfqsadsavtsifnrslyvfkqlrrefdqafqsytsfgtdvtepfpslsepaya 239		
Qy 270 DLEQCDIPNFPQLFCNFSVIYSVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329		
Db 240 daepsualpnvqilcnlsfsyqsvseklittratedppkqkdsnqggpiskilpeq 299		
Qy 330 DRGLCGELDQNLSCFRFHEKCKQCAHLSEDCPDVPAHTELDEATRLVNVSNQOYQOI 389		
Db 300 drgsdglgqnlsdcvnrkrcqcdqylsddcpnvpeylrelnealrlvsrsnqqydv 359		
Qy 390 LQMTKRKLEDYALVERMRGQFGWVSELANQAPETEIFNSIQVVPRI--HEGNSIKQDE 447		
Db 360 vqmtgyhledtllmekmreqfgwvselayspgaedifnpgvmvalsahegssgqdd 419		
Qy 448 TMTDLSILPSSNFTLKIPLEESAESSNFTGYGVAKALQHFKEHFTW 495		
Db 420 tvvps-silpssnftlsplesksagnanfidhvvekvqlghfkehftw 466		
RESULT 7		
AY30788		
ID AAY30788 standard; Protein: 450 AA.		
XX		
AC AAY30788;		
XX		
DT 23-NOV-1999 (first entry)		
XX		
DE	Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.	

XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	neuropyschiatric disorder; early-onset autosomal dominant myopia;	
KW	schizophrenia; splice variant.	
XX		
OS	Cavia cobaya.	
XX		
PN	WO9947535-A1.	
XX		
XX	23-SEP-1999.	
PD		
XX		
PF	16-MAR-1999; 99WO-US05606.	
XX		
XX	16-MAR-1998; 98US-0078044.	
PR	05-JUN-1998; 98US-0088312.	
PR	28-OCT-1998; 98US-0106056.	
PR	22-JAN-1999; 99US-0236134.	
XX		
XX	(MILL-) MILLENNIUM PHARM INC.	
PA	(REGC ) UNIV CALIFORNIA.	
PI	Chen H, Freimer NB;	
XX		
DR	WPI; 1999-562047/47.	
DR	N-PSDB; AA210756.	
XX		
XX	New HKNG1 polynucleotides useful in diagnosis and treatment of	
PT	neuropyschiatric disorders, e.g. bipolar affective disorders and	
PT	schizophrenia -	
XX		
PS	Claim 1; Fig 8A-B; 205pp; English.	
XX		
XX	The present sequence is encoded by a guinea pig HKNG1 ortholog	
CC	gphkng1815 splice variant gphkng7c. HKNG1 (Hong Kong new gene 1)	
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1	
CC	polynucleotides are useful to identify compounds modulating HKNG1	
CC	gene expression or HKNG1 polypeptide expression/activity. Compounds	
CC	inhibiting or enhancing HKNG1 gene expression or activity in	
CC	individuals can then be administered therapeutically to treat	
CC	HKNG1-mediated disorders, especially neuropyschiatric disorders	
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	in gene therapy techniques to treat such disorders. They are also useful	
CC	in diagnosis to identify individuals having, or at risk of developing,	
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	mutations especially result in the production of a protein with a	
CC	different sequence to the human full-length HKNG1 polypeptide or	
CC	splice variant sequences, especially the substitution of a lysine for	
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	useful in gene mapping, to produce probes or primers to identify	
CC	similar sequences (e.g. mutants or sequences from different species)	
CC	and to produce transgenic animals.	
XX		
SQ	Sequence 450 AA;	
Query Match		
Best Local Similarity 59.2%; Score 1558; DB 20; Length 450;		
Matches 299; Conservative 62; Mismatches 87; Indels 20; Gaps 3;		
QY	30 MKPPLVLFVILWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEEVKALTGTQKM 89	
Db	1 mkpllmfvcvllwkdcapcwkdktaisenanfseageldvgevkialigtkmqk 60	
QY	90 IMMERKEHTNLMSTLKKCRKQBALKLLNEVQHEELERESLADSGECRSL 149	
Db	61 immerreeehsklmktlkkckeekgealklmnevhehleeeelslcqvsldswdecrac 120	
QY	150 ENNCRIYTCQPSWSVKNKIERFRKIYQFLFPHEHNEKDLPISEKLIEDAQLTQM 209	
Db	121 esncmfdttcqpawsvknm-----endrgpvskgvteadaqvshi 163	
QY	210 EDVFSOLTVDVNSLFRNSFVFMQOEFDQTFQSHFISDTDLTEPYFFAFSPKPTKA 269	

Db	164 ehvflsldadvtslfnrslyvfkqirrefdqafqsgtvdcepfiffpslskepayra 223	
QY	270 DLEQCWDIPNFQFCNFSYIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 329	
Db	224 daepsaipnfvqlcnlsfsvygsvekittlratdppkdkdsnggppiskilpeq 283	
QY	330 DRGLCGELDONLSRCFKFHEKCKCQAHLSDECDPVPALHTEDEAIRLVNVSNOQYQGI 389	
Db	284 drgsdglglnsdcnvfrkrcqkcdqylsddcpnvpeylrelnearlrvsrnqgydqv 343	
QY	390 LOMTRKHLEDTAYLVEKMRGQFCWSELANQAPETELIFNSIOVVPRI--HEGNISKODE 447	
Db	344 vqmtgynleotllmekmreqfgvwselayspgaedifnpvkvmlsahnegnsdqqd 403	
QY	448 TMTDLSILPSSNFTLKIPLEESAESSNFYGVVAKALQHFKEHFKTW 495	
Db	404 tvvps-sllpsnftlsspleksagnanfidhvvkvqlqhfkehfkwtw 450	
RESULT 8		
AA30789		
ID	AA30789 standard; Protein; 374 AA.	
XX		
AC	AA30789;	
XX		
DT	23-NOV-1999 (first entry)	
XX		
DE	Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7c.	
XX		
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	neuropyschiatric disorder; early-onset autosomal dominant myopia;	
XX	schizophrenia; splice variant.	
OS	Cavia cobaya.	
PN	WO9947535-A1.	
XX		
PD	23-SEP-1999.	
XX		
PF	16-MAR-1999; 99WO-US05606.	
XX		
PR	16-MAR-1998; 98US-0078044.	
PR	05-JUN-1998; 98US-0088312.	
PR	28-OCT-1998; 98US-0106056.	
PR	22-JAN-1999; 99US-0236134.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
PA	(REGC ) UNIV CALIFORNIA.	
XX		
PI	Chen H, Freimer NB;	
XX		
DR	WPI; 1999-562047/47.	
DR	N-PSDB; AA210757.	
XX		
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of	
PT	neuropyschiatric disorders, e.g. bipolar affective disorders and	
PT	schizophrenia -	
XX		
PS	Claim 1; Fig 9A-B; 205pp; English.	
XX		
CC	The present sequence is encoded by a guinea pig HKNG1 ortholog	
CC	gphkng1815 splice variant gphkng7c. HKNG1 (Hong Kong new gene 1)	
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1	
CC	polynucleotides are useful to identify compounds modulating HKNG1	
CC	gene expression or HKNG1 polypeptide expression/activity. Compounds	
CC	inhibiting or enhancing HKNG1 gene expression or activity in	
CC	individuals can then be administered therapeutically to treat	
CC	HKNG1-mediated disorders, especially neuropyschiatric disorders	
CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	in gene therapy techniques to treat such disorders. They are also useful	
CC	in diagnosis to identify individuals having, or at risk of developing,	
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	mutations especially result in the production of a protein with a	
CC	different sequence to the human full-length HKNG1 polypeptide or	
CC	splice variant sequences, especially the substitution of a lysine for	
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	useful in gene mapping, to produce probes or primers to identify	
CC	similar sequences (e.g. mutants or sequences from different species)	
CC	and to produce transgenic animals.	
XX		

CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 374 AA;

Query Match 48.5%; Score 1277; DB 20; Length 374;  
 Best Local Similarity 54.5%; Pred. No. 2.4e-97;  
 Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

QY 30 MKPPLLVIIVCLLWKDHCAPTWKDKTAISENLKSFSEVGEIDADAEVKKALTIQOMK 89  
 DB 1 mkpllmfpvcilwkchcaptwkdktaisenansseageidvdgcvkialigkqm 60  
 QY 90 IMMERKEHTNLMSTLKKREEKQEAALKLINEVQHEERLCRESLADSWGECRSCL 149  
 DB 61 immerreeehsklmktlkckeeqalklmnevheheeslcqysladsdwdetrac 120  
 QY 150 ENNCMRIYTTCPQSSWSVKNKIERFRKIIYQFLPPFHEDNEKDLPISEKLIIEEDAQLTQM 209  
 DB 121 esncmrftttcqpawssvkn----- 140  
 QY 210 EDVFSQLTVDVNSLNFNSFNVFRMQQEFQDTFQSHFISDTLTPYFFPAFSKEPMTKA 269  
 DB 141 -----mepayra 147  
 QY 270 DLEQCWDIPNFQFCNFSYIYESVSETITKMLKATIEDLPKQKAPDHGGLSKMLPGQ 329  
 DB 148 daepsaipaivnqilcnlsfsyvsqsklittlratdppkqkdsngqppiskilpeq 207  
 QY 330 DRGLCGELDONLSRCFHEKCKQCAHLSEDCPDVPAHTELDATRLVNSVNOQYQGI 389  
 DB 208 drgsdglglnsdvcnfrkrcqkcdylsdcpnvpelyrelnearlsvrsnqydvq 267  
 QY 390 LQMTKRHLEDYALVEKMRGFGWVSELANOAPETEIFNSIOVPR--HEGNSKQDE 447  
 DB 268 vqmtqyhltdtlmexmrqfgwvselayqspgaedfnpvkmvmsahnegnsqdd 327  
 QY 448 TMTFTLSILSSNPTLPILESASNSNFTGYVYVAKALQHFKEHKTW 495  
 DB 328 tvtps-sllpsnsntflspleksagnanfidhvvekvlgfhkehtkw 374

RESULT 9  
 AAY30790  
 ID AAY30790 standard; Protein; 373 AA.  
 XX  
 AC AAY30790;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX

Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 XX

Cavia cobaya.  
 OS  
 XX  
 PN WO9947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-US05606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 XX  
 PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.  
 PR 22-JAN-1999; 99US-0236134.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chen H, Freimer NB;  
 XX  
 DR WPI: 1999-562047/47.  
 DR N-PSDB: AAZ10758.  
 XX  
 PT New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia  
 XX  
 PS Claim 1; Fig 10A-B; 205pp; English.  
 XX  
 CC The present sequence is encoded by a guinea pig HKNG1 ortholog  
 CC gphkng1815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)  
 CC is a gene associated with bipolar affective disorder (BAD). HKNG1  
 CC polynucleotides are useful to identify compounds modulating HKNG1  
 CC gene expression or HKNG1 polypeptide expression/activity. Compounds  
 CC inhibiting or enhancing HKNG1 gene expression or activity in  
 CC individuals can then be administered therapeutically to treat  
 CC HKNG1-mediated disorders, especially neuropsychiatric disorders  
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 373 AA;

Query Match 48.3%; Score 1272.5; DB 20; Length 373;  
 Best Local Similarity 54.5%; Pred. No. 5.7e-97;  
 Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 30 MKPPLLVIIVCLLWKDHCAPTWKDKTAISENLKSFSEVGEIDADAEVKKALTIQOMK 89  
 DB 1 mkpllmfpvcilwkchcaptwkdktaisenansseageidvdgcvkialigkqm 60  
 QY 90 IMMERKEHTNLMSTLKKREEKQEAALKLINEVQHEERLCRESLADSWGECRSCL 149  
 DB 61 immerreeehsklmktlkckeeqalklmnevheheeslcqysladsdwdetrac 120  
 QY 150 ENNCMRIYTTCPQSSWSVKNKIERFRKIIYQFLPPFHEDNEKDLPISEKLIIEEDAQLTQM 209  
 DB 121 esncmrftttcqpawssvkn----- 140  
 QY 210 EDVFSQLTVDVNSLNFNSFNVFRMQQEFQDTFQSHFISDTLTPYFFPAFSKEPMTKA 269  
 DB 141 -----mepayra 146  
 QY 270 DLEQCWDIPNFQFCNFSYIYESVSETITKMLKATIEDLPKQKAPDHGGLSKMLPGQ 329  
 DB 147 daepsaipaivnqilcnlsfsyvsqsklittlratdppkqkdsngqppiskilpeq 206  
 QY 330 DRGLCGELDONLSRCFHEKCKQCAHLSEDCPDVPAHTELDATRLVNSVNOQYQGI 389  
 DB 207 drgsdglglnsdvcnfrkrcqkcdylsdcpnvpelyrelnearlsvrsnqydvq 266  
 QY 390 LQMTKRHLEDYALVEKMRGFGWVSELANOAPETEIFNSIOVPR--HEGNSKQDE 447  
 DB 267 vqmtqyhltdtlmexmrqfgwvselayqspgaedfnpvkmvmsahnegnsqdd 326

QY	448	TMTDLSILPSSNFTLKIDLESAESSNFYGVVAKALQHFKEHFKTW	495
Db	327	tvvps-sllpssnftlsspleksagnanfidhvvkvlqhfekfktw	373
RESULT 10			
AAU28048	AAU28048 standard; Protein; 521 AA.		
XX	AC	AAU28048;	
XX	DT	23-NOV-1999 (first entry)	
XX	DE	Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.	
XX	KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
XX	KW	neuropyschiatric disorder; early-onset autosomal dominant myopia;	
XX	KW	schizophrenia; splice variant.	
OS	XX	Homo sapiens.	
PN	XX	WO9947535-A1.	
XX	PD	23-SEP-1999.	
XX	PF	16-MAR-1999; 99WO-US05606.	
XX	PR	16-MAR-1998; 98US-0078044.	
XX	PR	05-JUN-1998; 98US-0088312.	
XX	PR	28-OCT-1998; 98US-0106056.	
XX	PR	22-JAN-1999; 99US-0236134.	
XX	PA	(MILL-) MILLENNIUM PHARM INC.	
XX	PA	(REGC) UNIV CALIFORNIA.	
XX	PI	Chen H, Freimer NB;	
XX	XX	WPI; 1999-562047/47.	
DR	XX	N-PSDB; AAU28048.	
XX	PT	New HKNG1 polynucleotides useful in diagnosis and treatment of	
XX	PT	neuropyschiatric disorders, e.g. bipolar affective disorders and	
XX	PT	schizophrenia	
XX	PS	Claim 1; Fig 18A-B; 205pp; English.	
XX	CC	The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)	
XX	CC	splice variant. HKNG1 is a gene associated with bipolar affective	
XX	CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compounds	
XX	CC	modulating HKNG1 gene expression or HKNG1 polypeptide expression/	
XX	CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression or	
XX	CC	activity in individuals can then be administered therapeutically to	
XX	CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders	
XX	CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
XX	CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
XX	CC	in gene therapy techniques to treat such disorders. They are also useful	
XX	CC	in diagnosis to identify individuals having, or at risk of developing,	
XX	CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
XX	CC	mutations especially result in the production of a protein with a	
XX	CC	different sequence to the human full-length HKNG1 polypeptide or	
XX	CC	splice variant sequences, especially the substitution of a lysine for	
XX	CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
XX	CC	useful in gene mapping, to produce probes or primers to identify	
XX	CC	similar sequences (e.g. mutants or sequences from different species)	
XX	CC	and to produce transgenic animals.	
XX	XX	Sequence 521 AA;	
SQ	Query Match		
	Best Local Similarity	35.2%; Score 925.5; DB 20;	Length 521;
	Matches 181; Conservative	77.0%; Pred. No. 4.8e-68;	
		7; Mismatches 16;	Indels 31; Gaps 3;

QY	1	MKIKAEKNEGSRSSWQLHWGDIANNNGNKKPPLLVIIVCLLWLKDSHCAPTWKDKTATS	60
Db	94	mkikaeknegpsrswqlhwgdiannngnkkppllviivclllwlkdschcaptwkdktaia	153
QY	61	ENLKSPSEVGEIDADAEVKKALTIQKIMMERKEKHTNLMSTLKKCREKQKQKALKLL	120
Db	154	enlkspsevgeidadeevkkaitgikqkimmerkehtnlnmstlikkcreekqkalkll	213
QY	121	NEVQEHLEEEERLCRESLADSWGECRSCLNENCMRIYTTTCQPSWSSVKNKI--ERPFERKI	178
Db	214	nevqehleeeerlcrsladswgecrsclennmriytttcqpswssvknklltteafqrc	273
QY	179	YQFLPFPHEDNEKDLPISEKLIIEEDAQLTQMEDVFSQLT---VDVNSLFNRSPNV	230
Db	274	y-----lgrtedcvgnltricqdvsnfmknknv	302
RESULT 11			
AAU28048	AAU28048 standard; Protein; 449 AA.		
XX	AC	AAU28048;	
XX	DT	18-DEC-2001 (first entry)	
XX	DE	Novel human secretory protein, Seq ID No 217.	
XX	KW	Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;	
XX	KW	ischaemia-reperfusion injury; haematopoesis; cancer; neuropathy;	
XX	KW	transgenic animal; Alzheimer's disease; Parkinson's disease; burn;	
XX	KW	amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;	
XX	KW	ulcer; osteoporosis; bone degenerative disorder; periodontal disease;	
XX	KW	gut protection; lung; liver fibrosis; immune deficiency; infection;	
XX	KW	severe combined immunodeficiency; SCID; autoimmune disorder; allergy;	
XX	KW	multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;	
XX	KW	fertility; analgesic; pain; antigen.	
OS	XX	Homo sapiens.	
PN	XX	WO200166689-A2.	
XX	PD	13-SEP-2001.	
XX	PF	05-MAR-2001; 2001WO-US04942.	
XX	PR	07-MAR-2000; 2000US-0519705.	
XX	PR	19-MAY-2000; 2000US-0574454.	
XX	PR	17-JUN-2000; 2000US-0596193.	
XX	PR	14-JUL-2000; 2000US-0618847.	
XX	PR	19-SEP-2000; 2000US-0665363.	
XX	PR	20-OCT-2000; 2000US-0693267.	
XX	PA	(HYSE-) HYSEQ INC.	
XX	PI	Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;	
XX	PI	Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;	
XX	DR	WPI; 2001-589934/66.	
XX	DR	N-PSDB; AAS44948.	
XX	PT	Novel polypeptides and nucleic acids obtained from cDNA libraries	
XX	PT	prepared from various human tissues, for diagnosis and treatment of	
XX	PT	cancer, neurological, inflammatory, and autoimmune disorders -	
XX	PS	Example 3; SEQ ID No 217; 107pp; English.	
XX	CC	The invention relates to novel isolated human secreted polypeptides (I)	
XX	CC	and polynucleotides (II). (i) and (ii) are useful for treating	
XX	CC	inflammatory conditions such as arthritis, nephritis, Crohn's disease,	
XX	CC	ischaemia-reperfusion injury, shock, sepsis, immune responses, and is	
XX	CC	involved in increasing haematopoesis, stem cell survival, bone growth	
XX	CC	and remodeling. (I), (II) and modulators of (II) are useful for	
XX	CC	prophylaxis or treatment of one or more cancers. (II) is also useful for	

CC creating transgenic animals useful for studying the in vivo activities of  
 CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (1) induces the proliferation of neural cells and regeneration of nerve  
 CC and brain tissue and is useful for the treatment of central and  
 CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
 CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
 CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic  
 CC activity, regulation of haematopoiesis and is useful for treating myeloid  
 CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
 CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
 CC tissue growth, and in tissue repair, healing of burns, incisions,  
 CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
 CC disorders, or periodontal disease. Furthermore, (1) is also useful for  
 CC gut protection or regeneration and treatment of lung or liver fibrosis,  
 CC reperfusion injury in various tissues, various immune deficiencies and  
 CC disorders including severe combined immunodeficiency (SCID), bacterial or  
 CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
 CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
 CC reactions and conditions, such as asthma or other respiratory problems.  
 CC In addition, (1) affects biorhythms or circadian cycles of rhythms,  
 CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
 CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
 CC analgesic effects or other pain reducing effects, immunoglobulin like  
 CC activity and can act as an antigen in a vaccine composition to raise an  
 CC immune response. AAU2820-AAU28395 represent novel human secreted protein  
 CC amino acid sequences of the invention.

Sequence 448 AA;

Query Match 16.5%; Score 434; DB 22; Length 449;  
 Best Local Similarity 25.8%; Pred. No. 1.9e-27;  
 Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

Qy 30 MKPPLVFLVCLLWKDSCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 83  
 Db 1 mmktllivgll-----twesgvgldqtvsdnelqemsnqgskynkeiqnavn 51  
 Qy 84 GIKQMKIMMERKEHTNLMSTLKKREKQKALLLNEVQHEERLCRESLADSWG 143  
 Db 52 gvkqikltiektneerktllnleakkkedalnretsetkklpgvcnetmmalwe 111  
 Qy 144 ECRSCLNENNRITY-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEE 202  
 Db 112 eckpcikqctcmkfyarvcrgsglvgrgleeflngsspfywmngdr-----idsllen 165  
 Qy 203 DAQLTQMEDV----FSQLTVDVNSLNFNSFNVFRQMOQEFDTQFSHFIS-DTDLTEPYF 257  
 Db 166 drqqtthldvmqdhfrassildefqdrf-----ftrepqdy--hylpfslphrrphf 218  
 Qy 258 FPAFSKEPTKADLE-QCWDIPNFFQLFCNFSVSIYESVSITITKMLKAIEDLPKQDKAP 316  
 Db 219 f--fpksrivrslmpfspyepnlhamqpflemihe-----qqamdihfspaf 267  
 Qy 317 DHGGLISKMLPGQDRGLCGELDQNLSCFKFEKCKOKCOAHLSEDC----PDVPALHTEL 372  
 Db 268 qhpptefiregddrtvcrelhrstgclrmkdcokreilsvdcstnmpsqaklrrl 327  
 Qy 373 DEAIRLVNSVQQYQGIQLQTRKLEDTAYLVERMRGQFGWVSELANOAPETEIFNSIQ 432  
 Db 328 deslqvaerlkrkynellksyqwmntssllqelneqfnvwsrlantlqgedqyirvt 387  
 Qy 433 VVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLAESAESNFIYGVYAKALQHF-KEH 491  
 Db 388 tvas-htsdsdpsvgvtevvvklfdspitvtvpvevskrpkmetvaekalqeyrkkh 446

RESULT 12

AA11704

ID AA11704 standard; Protein; 448 AA.

XX

AC AA11704;

XX

DT 20-JUN-1991 (first entry)  
 XX Cytolysis Inhibitor.  
 DE Cytolysis Inhibitor.  
 KW cytotoxic inhibitor; perforin; immunological effector molecule;  
 XX infertility.  
 OS Homo sapiens.  
 FH Key Location/Qualifiers  
 FT Peptide 1..21  
 FT Protein /label= signal peptide  
 FT Protein 22..226  
 FT Protein /label= A-chain  
 FT Protein 227..448  
 FT Protein /label= B-chain  
 XX DE3933850-A.  
 XX 18-APR-1991.  
 XX 06-OCT-1989; 89DE-3933850.  
 XX 06-OCT-1989; 89DE-3933850.  
 XX (SCHD ) SCHERING AG.  
 XX Tschopp J, Jenne D;  
 XX WPT; 1991-118338/17.  
 XX DNA sequence coding for cytotoxic inhibitor - is strong inhibitor  
 XX of terminal complement protein, eg perforin secreted by killer  
 XX cells  
 XX Claim 13; Page 9; 15pp; German.  
 XX This cytotoxic inhibitor is encoded by a 1.7kb BamHI-KpnI fragment  
 XX isolated from a liver-specific cDNA library. It is a blood plasma  
 XX component that inhibits immunological effector molecules. It is used  
 XX for systemic or local treatment of inflammatory or autoimmune  
 XX diseases mediated by complement or killer cells. It can also be  
 XX used for detoxification of membrane-active and cytolytic proteins  
 XX released by bacteria, fungi and insect venoms. The protein has a  
 XX further use in the treatment of infertility caused by deficiency of  
 XX cytotoxic inhibitor. Monoclonal antibodies directed against the used  
 XX cytotoxic inhibitor are also covered by the invention. They are used  
 XX to isolate or quantify the natural protein in human plasma.  
 XX See also AAQ11501 and AAQ11502.  
 XX Sequence 448 AA;

Query Match 16.4%; Score 432; DB 12; Length 448;  
 Best Local Similarity 25.8%; Pred. No. 2.7e-27;  
 Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

Qy 34 LLFVIVCLLWKDSCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIKQ 87  
 Db 4 lllfvigll-----twesgvgldqtvsdnelqemsnqgskynkeiqnavngvkq 54  
 Qy 88 MKIMMERKEHTNLMSTLKKREKQKALLLNEVQHEERLCRESLADSWGECRS 147  
 Db 55 iktliektneerktllnleakkkedalnretsetkklpgvcnetmmalweeckp 114  
 Qy 148 CLENNENNRITY-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLEEADQL 206  
 Db 115 clktcmkfyarvcrgsglvgrgleeflngsspfywmngdr-----idsllendrqq 168  
 Qy 207 TOMEDV----FSQLTVDVNSLNFNSFNVFRQMOQEFDTQFSHFIS-DTDLTEPYFPAP 261  
 Db 169 thldvmqdhfrassildefqdrf-----ftrepqdy--hylpfslphrrphf--f 219





```
XX Key Location/Qualifiers
FH Peptide 1..22
FT /label= signal_peptide
FT Protein 23..225
FT /label= mature_clusterin_alpha_chain
FT Protein 226..446
FT /label= mature_clusterin_beta_chain
FT Region 74..80
FT /label= nuclear_localisation_signal
FT Modified-site 86..88
FT /label= potential_glycosylation_site
FT Modified-site 103..105
FT /label= potential_glycosylation_site
FT Modified-site 145..147
FT /label= potential_glycosylation_site
FT Modified-site 225..226
FT /label= interchain_cleavage_site
FT /note= "leads to the creation of alpha and beta chains"
FT Modified-site 290..292
FT Modified-site 316..318
FT /label= potential_glycosylation_site
FT Modified-site 353..355
FT /label= potential_glycosylation_site
FT Modified-site 373..375
FT /label= potential_glycosylation_site
XX
XX W0200034469-A1.
XX
XX 15-JUN-2000.
XX
XX 10-DEC-1999; 99WO-US29262.
XX
XX 11-DEC-1998; 98US-O111856.
XX
XX (UYNY ) UNIV NEW YORK STATE RES FOUND.
XX
XX Millis AJT;
XX
XX WPI; 2000-431300/37.
XX
XX Clusterin and gp38k-related peptide capable of altering cell migration
XX useful for treating atherosclerosis, cancer and stenosis following
XX vascular trauma or disease
XX
XX Disclosure; Fig 1; 43pp; English.
XX
XX The present sequence is the protein sequence of porcine clusterin.
XX Clusterin (also known as complement cytolytic inhibitor, sulfated
XX glycoprotein-2, testosterone repressed prostate message-2, SP-40, 40 and
XX ApoJ) is essential for the migration of vascular smooth muscle cells
XX (VSMC). The gene and protein can, therefore, be used to promote wound
XX healing, angiogenesis and vasculogenesis, in the treatment of stenosis
XX following vascular trauma or disease and to treat atherosclerosis, and
XX antisense sequences can be used to treat cancer, as angiogenesis is
XX vital for tumour survival.
XX
XX Sequence 446 AA;

Query Match 15.2%; Score 400.5; DB 21; Length 446;
Best Local Similarity 24.4%; Pred. No. 1.1e-24;
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

Oy 34 LLVFVCLLWLKDSHCAPTWK-----DKTAISENLKSFSEVEIDAEVKKALTGK 86
Db 4 llllvgl11-----twngpwlvgdkalsdkelqemstegskynkeiknalkev 54
Oy 87 QMKIMWERKEHTNLMSTLKKCREKQKALKLLNEVQHELEERLCRESLADSWGECR 146
Db 55 qiktliedqneerksllsleakkkkdaIndtrdtetk1kgsgglcnetmmalweeck 114
```

```
OY 147 SCLNNCMRIYT--TCOPSSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISKEKLEBDAQ 205
Db 115 pelkqtcmkfyarvcrgsglvgqleeflnqspfyfwingdr-----idslmemdrq 168
OY 206 LTOMEDVFSOLTVDVNSLFRNSFVFRMQOEFDQTFQSHFIS-----DIDLTEPY----- 256
Db 169 qshvmdi-----medsforasni-----mdelfqdrfndrepfdtqfsgfgsshr 214
OY 257 ---FPAPAFSKPEWTKADLEOCWDIPNFFOLFNCFSVSIVESVETITTKMLKATIEDLPKQD 313
Db 215 gslfnpkrsfarnimpflftdi-nyndmfqpf-----fdmihqagqamdahhr1pyh- 268
OY 314 KAPDHGGLISKMLP--CQDRGLAGELDONLSRCFKFHEKQCKQAOHLSEDC-----PDVPA 367
Db 269 -fpeag-----vpensndravckeirhnstgclrmkdqckecreilsvdcsaanssqmq 321
OY 368 LHTELDEAIRLVNVSNOQYQIQLQMKRKHLEDTAYLVKMRGQFGWVSELANQAPETEI 427
Db 322 lrqelytslqmaekfsklydqllysqgkmlntsslikqineqfswsqianitqnddry 381
OY 428 FNSIQVVPRIHEGNISKODETMTDLSILPSSNFTLPILEESAESSNFYGVVAKALQH 487
Db 382 ylvttv-nshgspsvpsgltkvvvkldfyspyitlilpqevs--dpkfmvtvaeaalq 438
OY 488 FKE 490
Db 439 yrq 441

RESULT 15
AAE03764
ID AAE03764 standard; Protein; 363 AA.
XX
XX AAE03764;
XX
XX 07-AUG-2001 (first entry)
XX
XX Human gene 1 encoded secreted protein HOF0C33, SEQ ID NO:34.
XX
XX Human; secreted protein; proliferative disorder; cancer; tumour;
XX foetal abnormality; developmental abnormality; haematopoietic disorder;
XX immune system disorder; AIDS; autoimmune disease; rheumatoid arthritis;
XX inflammation; allergy; neurological disorder; Alzheimer's disease;
XX Parkinson's disease; cognitive disorder; schizophrenia; asthma;
XX skin disorder; psoriasis; sepsis; diabetes; atherosclerosis;
XX cardiovascular disorder; angiogenic disorder; kidney disorder;
XX gastrointestinal disorder; pregnancy-related disorder; gene therapy;
XX endocrine disorder; infection; wound healing; vulneryary;
XX cell culture; chemotaxis; food additive;
XX binding partner identification.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
FH Peptide 1..21
FT /label= signal_peptide
FT Protein 22..363
FT /note= "Mature secreted protein"
XX
XX W0200132837-A1.
XX
XX 10-MAY-2001.
XX
XX 17-OCT-2000; 2000WO-US28664.
XX
XX 02-NOV-1999; 99US-0163085.
XX 17-DEC-1999; 99US-0172411.
XX
XX (HUMA-) HUMAN GENOME SCI INC.
XX
XX Fiscella M, Wei P, Lafleur DW, Olsen HS, Baker K, Ebner R;
XX Komatsoulis G, Rosen CA, Ruben SM, Duan RD, Young PE, Florence KA;
XX Moore PA, Birse CE, Ni J, Soppet DR, Shi Y;
```





Db 241 TFSHFISDITLTPYFFPAFSKEPTWKADLEQCWDIPNFFQLFCNFSVIIVSETIT 300  
QY 301 KMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKFHEKCKQCAHLSE 360  
Db 301 KMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKFHEKCKQCAHLSE 360  
QY 361 DCPDVPALHTDELATRLVNVSNQOYGOILOMTRKHLEDATAYLVKMRGQFGWVSELNQ 420  
Db 361 DCPDVPALHTDELATRLVNVSNQOYGOILOMTRKHLEDATAYLVKMRGQFGWVSELNQ 420  
QY 421 APETEIFNSIOVVPRIHEGNTSKQDETMTDLSILPSSNFTLKIPLESAESSNFIGYV 480  
Db 421 APETEIFNSIOVVPRIHEGNTSKQDETMTDLSILPSSNFTLKIPLESAESSNFIGYV 480  
QY 481 VAKALQHFKHEFKTW 495  
Db 481 VAKALQHFKHEFKTW 495

## RESULT 2

US-09-268-992-64

; Sequence 64, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; CURRENT FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; EARLIER FILING DATE: 1998-03-16

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 64

; LENGTH: 466

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-268-992-64

Query Match 93.6%; Score 2464; DB 4; Length 466;  
Best Local Similarity 100.0%; Pred. No. 2.4e-228; Mismatches 0; Indels 0; Gaps 0;  
Matches 466; Conservative 0;

QY 30 MKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEDADAEVKKALTGKQMK 89  
Db 1 MKPPLLVFVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEDADAEVKKALTGKQMK 60  
QY 90 IMMERKEHTNLMSTYLLKKREKQKALKNLNEVQHLSEERLCRESLADSWGECRSL 149  
Db 61 IMMERKEHTNLMSTYLLKKREKQKALKNLNEVQHLSEERLCRESLADSWGECRSL 120  
QY 150 ENNCRIYTTCPQSWSSVSNKRIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 209  
Db 121 ENNCRIYTTCPQSWSSVSNKRIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQM 180  
QY 210 EDVFSOLTVDVNSLNRSENFVRMQQEDQTFQSHFISDITLTPYFFPAFSKEPTKA 269  
Db 181 EDVFSOLTVDVNSLNRSENFVRMQQEDQTFQSHFISDITLTPYFFPAFSKEPTKA 240  
QY 270 DLEQCWDIPNFFQLFCNFSVIIVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQ 329  
Db 241 DLEQCWDIPNFFQLFCNFSVIIVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQ 300

QY 330 DRGLCGELDNLSRCFKFHEKCKQCAHLSEDCPDVPALHTDELDEAIRLVNVSNQOYGOI 389  
Db 301 DRGLCGELDNLSRCFKFHEKCKQCAHLSEDCPDVPALHTDELDEAIRLVNVSNQOYGOI 360  
QY 390 LOWTRKHLEDATAYLVKMRGQFGWVSELANOAPETEIFNSIQVVPRIHEGNTSKODETM 449  
Db 361 LOWTRKHLEDATAYLVKMRGQFGWVSELANOAPETEIFNSIQVVPRIHEGNTSKODETM 420  
QY 450 MTDLSSLPSNFTLKIPLESAESSNFIGYVYAKALQHFKHEFKTW 495  
Db 421 MTDLSSLPSNFTLKIPLESAESSNFIGYVYAKALQHFKHEFKTW 466

## RESULT 3

US-09-268-992-51

; Sequence 51, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; CURRENT FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; EARLIER FILING DATE: 1998-03-16

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 51

; LENGTH: 446

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-268-992-51

Query Match 89.3%; Score 2350; DB 4; Length 446;  
Best Local Similarity 100.0%; Pred. No. 2e-217; Mismatches 0; Indels 0; Gaps 0;  
Matches 446; Conservative 0;

QY 50 APTWKDKTAISENLKSFSEVGEDADAEVKKALTGKQMKIMMERKEKHTNLMSTLKKC 109  
Db 1 APTWKDKTAISENLKSFSEVGEDADAEVKKALTGKQMKIMMERKEKHTNLMSTLKKC 60  
QY 110 REEKQKALKNLNEVQHLSEERLCRESLADSWGECRSCLENNCMRIYTTCPQSWSSVKN 169  
Db 61 REEKQKALKNLNEVQHLSEERLCRESLADSWGECRSCLENNCMRIYTTCPQSWSSVKN 120  
QY 170 KIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQMEDVFSOLTVDVNSLNRSEN 229  
Db 121 KIERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDAQLTQMEDVFSOLTVDVNSLNRSEN 180  
QY 230 VFRMQQEDQTFQSHFISDITLTPYFFPAFSKEPTKADLEQCWDIPNFFQLFCNFSV 289  
Db 181 VFRMQQEDQTFQSHFISDITLTPYFFPAFSKEPTKADLEQCWDIPNFFQLFCNFSV 240  
QY 290 STYESVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKPHE 349  
Db 241 STYESVSETITKMLKAIEDLPKQKADPHGGLSKMLPGQDRGLCGELDNLSRCFKPHE 300  
QY 350 KCKQCAHLSEDCPDVPALHTDELDEAIRLVNVSNQOYGOILOMTRKHLEDATAYLVKMRG 409  
Db 301 KCKQCAHLSEDCPDVPALHTDELDEAIRLVNVSNQOYGOILOMTRKHLEDATAYLVKMRG 360  
QY 410 QFGWVSELANOAPETEIFNSIQVVPRIHEGNTSKODETMMDLSILPSSNFTLKIPLEE 469  
Db 361 QFGWVSELANOAPETEIFNSIQVVPRIHEGNTSKODETMMDLSILPSSNFTLKIPLEE 420

Qy 470 SAESSNFIGYVAKALQHFKEHFKTW 495  
|||||  
Db 421 SAESSNFIGYVAKALQHFKEHFKTW 446

## RESULT 4

US-09-268-992-4

; Sequence 4, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; EARLIER FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 4

; LENGTH: 477

; TYPE: PRT

; ORGANISM: Homo sapiens

US-09-268-992-4

Query Match 88.7%; Score 2335; DB 4; Length 477;  
Best Local Similarity 91.2%; Pred. No. 6.3e-216;  
Matches 448; Conservative 2; Mismatches 1; Indels 40; Gaps 2;

Qy 25 NNSGNKPKPLLVFVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTG 84  
:|||||  
Db 7 SMSGNKPPLLVFVCLLWKDSHCAPTWKDKSAISENLKSFSEVGEIDAEVKKALTG 66  
:|||||  
Qy 85 IKQKIMMERKEK-----EHTNLMSTLKKCKREEKQKALKLLNEVQ 124  
:|||||  
Db 67 IKQKIMMERKEKANAPEITEIFNSIQVVPRIEHTNLMSTLKKCKREEKQKALKLLNEVQ 126  
:|||||  
Qy 125 EHLEEBERLCRESLADSGECRSCLENNCMRIYTTCCPSWSSVKNKIERFRKKIYQFLFP 184  
:|||||  
Db 127 EHLEEBERLCRESLADSGECRSCLENNCMRIYTTCCPSWSSVKNKIERFRKKIYQFLFP 186  
:|||||  
Qy 185 PHEDNEKDLPISEKLEEDAQLTOMEDVFSQLTVDVNSLFRNSFNVRMQOQERDOTFS 244  
:|||||  
Db 187 PHEDNEKDLPISEKLEEDAQLTOMEDVFSQLTVDVNSLFRNSFNVRMQOQERDOTFS 246  
:|||||  
Qy 245 HFISDTDLTEPYFFPAFSKEPTKADLEQCDWIPNFQFCNFSYIYESVSETITKMLK 304  
:|||||  
Db 247 HFISDTDLTEPYFFPAFSKEPTKADLEQCDWIPNFQFCNFSYIYESVSETITKMLK 306  
:|||||  
Qy 305 ATEDLPKQKAPDHPGGLISKMLPGDRLGCGELDONLSRCFKFHEKCKQKQAHLSQD 364  
:|||||  
Db 307 ATEDLPKQKAPDHPGGLISKMLPGDRLGCGELDONLSRCFKFHEKCKQKQAHLSQD 366  
:|||||  
Qy 365 VPALHTELDEARLVNSNOQYGOILQMTKRKLEDTAYLVEKMRGQFGVWSSELANQAPET 424  
:|||||  
Db 367 VPALHTELDEARLVNSNOQYGOILQMTKRKLEDTAYLVEKMRGQFGVWSSELANQAPET 419  
:|||||  
Qy 425 EIIFNSIQVVPRIHEGNISKQDETMMTDLISLPSSNFTLKIPLEESAESSNFIGYVAKA 484  
:|||||  
Db 420 -----HEGNISKQDETMMTDLISLPSSNFTLKIPLEESAESSNFIGYVAKA 466  
:|||||  
Qy 485 LQHFKEHFKTW 495  
|||||

Db 467 LQHFKEHFKTW 477

## RESULT 5

US-09-268-992-49

; Sequence 49, Application US/09268992

; Patent No. 6342351

; GENERAL INFORMATION:

; APPLICANT: Chen, H.

; APPLICANT: Freimer, N.

; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING

; FILE REFERENCE: 7853-138

; CURRENT APPLICATION NUMBER: US/09/268,992

; EARLIER FILING DATE: 1999-03-16

; EARLIER APPLICATION NUMBER: 09/236,134

; EARLIER FILING DATE: 1999-01-22

; EARLIER APPLICATION NUMBER: 60/106,056

; EARLIER FILING DATE: 1998-10-28

; EARLIER APPLICATION NUMBER: 60/088,312

; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044

; EARLIER FILING DATE: 1998-03-16

; NUMBER OF SEQ ID NOS: 84

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 49

; LENGTH: 465

; TYPE: PRT

; ORGANISM: Bos sp.

US-09-268-992-49

Query Match 66.7%; Score 1755.5; DB 4; Length 465;  
Best Local Similarity 70.8%; Pred. No. 2.8e-160;  
Matches 330; Conservative 58; Mismatches 77; Indels 1; Gaps 1;

Qy 30 MKPPLLVFIVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGKQMK 89  
:|||||  
Db 1 MKPPLLVFIVVLLQDCQCAPTKGRTSIREDPKGFSGAGEIDVDEEVKKALIGKQMK 60  
:|||||  
Qy 90 IMMERKEHEHNLMSTLKKCKREEKQKALKLLNEVOEHLEEBERLCRESLADSGECRSC 149  
:|||||  
Db 61 ILMERREHEHSLKMLTKKCKREEKQKALKLLNEVOEHLEEBERLCQVSLMGSDCKSCL 120  
:|||||  
Qy 150 ENNCMRIYTTCCPSWSSVKNKIERFRKKIYQFLFPFHEDNEKDLPISEKLEEDAQLTOM 209  
:|||||  
Db 121 ESDCMRFTTCCQSSWSSMKSTIERVFRKIYQFLFPFHEDDEKELPVGKFTTEEDVQLMQI 180  
:|||||  
Qy 210 EDVFSQLTVDVNSLFRNSFNVRMQOQERDOTFSQSHFISDTDLTEPYFFPAFSKEPTKA 269  
:|||||  
Db 181 ENVFSQLTVDVGFVLYNMSFHVFKQMQOQERDQAFQSYFNSDTSMEPYFFPAFSKEPAKKA 240  
:|||||  
Qy 270 DLEQCDWIPNFQFCNFSYIYESVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329  
:|||||  
Db 241 HPMQSWDIPSFQFQFCNFSYIYESVSETITKMLKAIEDLSKQDKSDSAHGSPSTTPVR 300  
:|||||  
Qy 330 DRGLCGELDONLSRCFKFHEKCKQKQAHLSQDQVDPALHTELDEARLVNSNOQYGOI 389  
:|||||  
Db 301 GRGLCGEPQNSSECLQFHARCQKQDYLWADCPAVPELYTKADEALELVNLSNOQYQAV 360  
:|||||  
Qy 390 LQMTKRKLEDTAYLVEKMRGQFGVWSSELANQAPETEIFNSIQVVPRIHEGNISKQDETMM 449  
:|||||  
Db 361 LQMTQHLEDITYLMEKMRGQFGVWTELSQTPGSESNFISFKVVPVGHGEGFSQDEKM 420  
:|||||  
Qy 450 MTDLSILPSSNFTLKIPLEESAESSNFTGYVAKALQHFKEHFKTW 495  
:|||||  
Db 421 I-DTSLPSSNFTLKIPLEESAESSDFTSYMLAKAVQHFKEHFKSW 465  
:|||||

## RESULT 6

US-09-268-992-39

; Sequence 39, Application US/09268992.

; Patent No. 6342351

```
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 39
; LENGTH: 466
; TYPE: PRT
; ORGANISM: Cavia sp.
; US-09-268-992-39

Query Match      62.3%; Score 1639; DB 4; Length 466;
Best Local Similarity 66.7%; Pred. No. 4.4e-149;
Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;

QY 30 MKPPLLVIIVCLLWLDKSHCAPTWKDKTAISENLKSPFSEVGEIDADBEVKKALTGKQMK 89
Db 1 MKPLPMPVCLLWLDKCHCAPTWKDKTAISENANSESEAGEIDVDGEVIALIGIKQMK 60
QY 90 IMMERKEHTNLMTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
Db 61 IMMERREEHSLKMLTKLKKCKEEKQKALKLMNEVHEHELEESLQCVSLADSWDECRACL 120
QY 150 ENCMRIYTTTCQPSWSSVKNKIERFRKIIYQFLFPFHEDNEKDLPISEKLIBEDAQITQM 209
Db 121 ESNCRFDTTCQPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163
QY 210 EDVFSQTLVDVNSLNFNRSENVFRQMOEQFDTQFQSHFISDITLDEPFFPAFSEKPMTKA 269
Db 180 EHVSQLSADVTSLEFNRLVFKQLREFDQAFQSVFTSGTDTTEPFFPSLSKEPAYRA 239
QY 270 DLQCDWIPNFFOLFCNFSVIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGQ 329
Db 240 DABPSWAIPNVFOLLNLSFSVQSVSEKLIITLRLATEDPPKQDKDSNQGPGISKILPEQ 299
QY 330 DRGLGCLDQNLRCFKFEKCKCOAHLSQEDCPDVPALHTEDEAIRLVNVSNOQYQGI 389
Db 300 DRSGDKLGNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 359
QY 390 LQWTRKHLEDYAVLVEKMRQFGWSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
Db 360 VQMTQVHLEDYTLLEKMRQFGWSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDDD 419
QY 448 TMTDLSILPSSNFTLKIPLEESAESSNFIYGVVAKALQHFKEHFTW 495
Db 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFTW 466

RESULT 7
US-09-268-992-41
; Sequence 41, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
```

```
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 41
; LENGTH: 450
; TYPE: PRT
; ORGANISM: Cavia sp.
; US-09-268-992-41

Query Match      59.2%; Score 1558; DB 4; Length 450;
Best Local Similarity 63.9%; Pred. No. 2.6e-141;
Matches 299; Conservative 62; Mismatches 87; Indels 20; Gaps 3;

QY 30 MKPPLLVIIVCLLWLDKSHCAPTWKDKTAISENLKSPFSEVGEIDADBEVKKALTGKQMK 89
Db 1 MKPLPMPVCLLWLDKCHCAPTWKDKTAISENANSESEAGEIDVDGEVIALIGIKQMK 60
QY 90 IMMERKEHTNLMTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149
Db 61 IMMERREEHSLKMLTKLKKCKEEKQKALKLMNEVHEHELEESLQCVSLADSWDECRACL 120
QY 150 ENCMRIYTTTCQPSWSSVKNKIERFRKIIYQFLFPFHEDNEKDLPISEKLIBEDAQITQM 209
Db 121 ESNCRFDTTCQPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163
QY 210 EDVFSQTLVDVNSLNFNRSENVFRQMOEQFDTQFQSHFISDITLDEPFFPAFSEKPMTKA 269
Db 164 EHVSQLSADVTSLEFNRLVFKQLREFDQAFQSVFTSGTDTTEPFFPSLSKEPAYRA 223
QY 270 DLQCDWIPNFFOLFCNFSVIYESVSETITKMLKAIEDLPKQKADPHGGLISKMLPGQ 329
Db 224 DABPSWAIPNVFOLLNLSFSVQSVSEKLIITLRLATEDPPKQDKDSNQGPGISKILPEQ 283
QY 330 DRGLGCLDQNLRCFKFEKCKCOAHLSQEDCPDVPALHTEDEAIRLVNVSNOQYQGI 389
Db 284 DRSGDKLGNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 343
QY 390 LQWTRKHLEDYAVLVEKMRQFGWSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447
Db 344 VQMTQVHLEDYTLLEKMRQFGWSELAYQSPGAEDIFNPVKVMVALSAHEGNSDDDD 403
QY 448 TMTDLSILPSSNFTLKIPLEESAESSNFIYGVVAKALQHFKEHFTW 495
Db 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFTW 450

RESULT 8
US-09-268-992-43
; Sequence 43, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
```

; EARLIER APPLICATION NUMBER: 60/078,044  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 43  
; LENGTH: 374  
; TYPE: PRG  
; ORGANISM: Cavia sp.  
US-09-268-992-43

Query Match 48.58; Score 1277; DB 4; Length 374;  
Best Local Similarity 54.58; Pred. No. 1.9e-114;  
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;  
  
QY 30 MKPPLVFIYVCLLWLDKSHCAPTKWKDTAISENLKSFSEVGEIDAEVKKALTGKQMK 89  
DB 1 MKPLLMFPVCLLWLDKSHCAPTKWKDTAISENANSESEAGEIDVGEVKIALIGIKQMK 60  
  
QY 90 IMMERKEHTNLMSTLKKCKEKEQKALKLLNEVQEHLEERLCRESLADSWGECRSL 149  
DB 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHEEESLCQVSLADSWDECRACL 120  
  
QY 150 ENNCRIYTTCPSSWSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEEDAQLTQM 209  
DB 121 ENSCMREDTTCOPAWSSVKN----- 140  
  
QY 210 EDVFSQLTVDVNSLFRSFRVFMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269  
DB 141 -----MPAYRA 146  
  
QY 270 DLEOCWDIPNFQFPCNFSVSIYESVETITKMLKAIEDLPKQKADPHGLSKMLPGQ 329  
DB 148 DAEPWAIPNVFQLLNLCSFSVQSVSEKLTITLRATEDPPKQKDSNOGGPISKILPEQ 207  
  
QY 330 DRGLCGELDONLSRCFKFHEKCKQKQAHLSDDCPDVPALHTEDEAIRLVNSNQYQOI 389  
DB 208 DRGSDGKLGQNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 267  
  
QY 390 LQMTKRKHELTAYLVERKMRGQGWVSELANQAPETEIFNSIQVVPRI--HEGNISKODE 447  
DB 268 VQMTQYHLEDITLLMEKMRQEGWVSELAYQSPCAEDIFNPVKVVALSAHEGNSDQDD 327  
  
QY 448 TMWTDLSILPSSNTFLKPILESASSNFYGVVYAKALQHFKEHFTW 495  
DB 328 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVVEKVLQHFKEHFTW 374

RESULT 9  
US-09-268-992-45  
; Sequence 45, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 45  
; LENGTH: 373  
; TYPE: PRG

; ORGANISM: Cavia sp.  
US-09-268-992-45

Query Match 48.38; Score 1272.5; DB 4; Length 373;  
Best Local Similarity 54.58; Pred. No. 5.1e-114;  
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 30 MKPPLVFIYVCLLWLDKSHCAPTKWKDTAISENLKSFSEVGEIDAEVKKALTGKQMK 89  
DB 1 MKPLLMFPVCLLWLDKSHCAPTKWKDTAISENANSESEAGEIDVGEVKIALIGIKQMK 60  
  
QY 90 IMMERKEHTNLMSTLKKCKEKEQKALKLLNEVQEHLEERLCRESLADSWGECRSL 149  
DB 61 IMMERREEHSLKMTLKKCKEKEQKALKLLNEVHEHEEESLCQVSLADSWDECRACL 120  
  
QY 150 ENNCRIYTTCPSSWSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEEDAQLTQM 209  
DB 121 ENSCMREDTTCOPAWSSVKN----- 140  
  
QY 210 EDVFSQLTVDVNSLFRSFRVFMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269  
DB 141 -----MPAYRA 146  
  
QY 270 DLEOCWDIPNFQFPCNFSVSIYESVETITKMLKAIEDLPKQKADPHGLSKMLPGQ 329  
DB 147 DAEPWAIPNVFQLLNLCSFSVQSVSEKLTITLRATEDPPKQKDSNOGGPISKILPEQ 206  
  
QY 330 DRGLCGELDONLSRCFKFHEKCKQKQAHLSDDCPDVPALHTEDEAIRLVNSNQYQOI 389  
DB 207 DRGSDGKLGQNLSDCVNFRKRCQKQDYLSDDCPNVPPELYRELNEALRLVRSRNOQYDQV 266  
  
QY 390 LQMTKRKHELTAYLVERKMRGQGWVSELANQAPETEIFNSIQVVPRI--HEGNISKODE 447  
DB 267 VQMTQYHLEDITLLMEKMRQEGWVSELAYQSPCAEDIFNPVKVVALSAHEGNSDQDD 326  
  
QY 448 TMWTDLSILPSSNTFLKPILESASSNFYGVVYAKALQHFKEHFTW 495  
DB 327 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVVEKVLQHFKEHFTW 373

RESULT 10  
US-09-268-992-67  
; Sequence 67, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 67  
; LENGTH: 208  
; TYPE: PRG  
; ORGANISM: Homo sapiens  
US-09-268-992-67

Query Match 35.08; Score 921; DB 4; Length 208;  
Best Local Similarity 99.48; Pred. No. 1.2e-80;  
Matches 170; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKTKAENEGPSRWOLHMGDIANNNGNKKPPLLVFIVCLLWKDSHCAPTWKDKTAIS 60  
DB 33 MKTKAENEGPSRWOLHMGDIANNNGNKKPPLLVFIVCLLWKDSHCAPTWKDKTAIS 92  
QY 61 ENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCREEKQKALKL 120  
DB 93 ENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCREEKQKALKL 152  
QY 121 NEVQHLLEERLCRESLADSWGECRSCLENNCMRIYTTCCPSWSSVKNKI 171  
DB 153 NEVQHLLEERLCRESLADSWGECRSCLENNCMRIYTTCCPSWSSVKNKL 203

## RESULT 11

US-08-685-576-1  
; Sequence 1, Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaibuchi, Kozo  
; APPLICANT: Iwamatsu, Akihiro  
; APPLICANT: Nakano, Takeshi  
; APPLICANT: Ito, Masaaki  
; APPLICANT: Takahashi, No. 5906819uaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109

COMPUTER READABLE FORM:  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 24-JUL-1996  
; APPLICATION NUMBER: US/08/685,576  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; FILING DATE: 20-NOV-1995  
; APPLICATION NUMBER: JP 7-325129  
; PRIOR APPLICATION DATA:  
; FILING DATE: 05-JAN-1996  
; APPLICATION NUMBER: JP 8-17150  
; PRIOR APPLICATION DATA:  
; FILING DATE: 26-APR-1996  
; APPLICATION NUMBER: JP 8-131206  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Bent, Stephen A.  
; REGISTRATION NUMBER: 29,768  
; REFERENCE/DOCKET NUMBER: 16887/843  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (202)672-5300  
; TELEFAX: (202)672-5399  
; TELEX: 904136

INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1388 amino acids  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
US-08-685-576-1  
Query Match 5.5%; Score 145; DB 2; Length 1388;  
Best Local Similarity 16.5%; Pred. No. 8.1e-05;  
Matches 94; Conservative 97; Mismatches 166; Indels 212; Gaps 20;

QY 54 KDKTAISENLKSFSEVEIDADEVKALGTGKQKIMMERKEKEHTNLMSTLKKCREEKQKALKL 104

DB 780 KQDVLNEDVRNLT-----LKIEQETQKRCLTQNDLQWQTOQVNTLAKMSEKQLKQENHLL 835  
QY 105 TLK-----KCREKQKALKLLNEVQHELEEE-----RLCR 135  
DB 836 ENKMSLEKQNAELRKERQDADGQMKELQDLQLEAEQVYSTLYKTQVRELKEECEEKTKLCK 895  
QY 136 E-----SLADSWGECRSCLENNCMRIYTTCCPSWSSVKNKIETFRFRKIYQFLFPFHED 189  
DB 896 ELQOKKQELQDE-----RDSLAAQLEITLTTRADSEQLARSAIEEQYSDEKE 942  
QY 190 E--KDLPISE-----KLIEDAQLTOMEDVFSQLTVDVNSLNSRFSNVFMQOQERDQ 240  
DB 943 KIMKELEIKEMMARHKQELTEKDATIASLEETNRTLTSDVANLANEKEELNKKLKEAQO 1002  
QY 241 TQSHFISDTDLTEPYFFPAFSKPEMTKADLEOCNDIPNFQFCNFSVSVSETIT 300  
DB 1003 LSR---LKDEISAAAKAQEFKQLLTERTLK-----TQAVN 1036  
QY 301 KMLKATE-----DLPKQDKAPDHGGLISKMLPGQDRGLCGEL-----DQNLSCF 345  
DB 1037 KLAEMNRKEPVKRGNDTQVRKEK-----ENRKLHMLKSEKREKLTQOMI 1082  
QY 346 KPEHKQKQCAHLSEDCPDVPAHHTDELDEAIRLVNVSQYQGIQLQMTKRHLE----- 398  
DB 1083 KYQKELNEMQAQIAEE---SQIRTELQMTLDSKDSIDIELRSQALHIGLDSSSIGSG 1138  
QY 399 -----DTAYLYVKMRG-----QFGWVSEL----- 417  
DB 1139 PGDEADGCPESRUEGWLSPVRNNTKFGVWYKVIYVSSKKILFYDSEQDKESNPYM 1198  
QY 418 -----ANQAPETETIIFNSIQVVPRI-----HEGNISKQDETMTMTDLSILPSSNFT 462  
DB 1199 VLDIDKLFHVRPVTQDVYRADAKEIPRFQILYANEGESKKEQE----- 1243  
QY 463 LKIPLEESAESNFI---GYVVAKALQHF 488  
DB 1244 --FPVEPVGEKSNYICHKGHEFIPTLYHF 1270

## RESULT 12

US-08-685-576-4  
; Sequence 4, Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaibuchi, Kozo  
; APPLICANT: Iwamatsu, Akihiro  
; APPLICANT: Nakano, Takeshi  
; APPLICANT: Ito, Masaaki  
; APPLICANT: Takahashi, No. 5906819uaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,576  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-325129  
; FILING DATE: 20-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-17150



;; FILING DATE: 05-JAN-1996  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: JP 8-131206  
;; FILING DATE: 26-APR-1996  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Bent, Stephen A.  
;; REGISTRATION NUMBER: 29,768  
;; REFERENCE/DOCKET NUMBER: 16887/843  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (202)672-5300  
;; TELEFAX: (202)672-5399  
;; TELEX: 904136  
;; INFORMATION FOR SEQ ID NO: 4:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 1388 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-685-576-4

Query Match 5.4%; Score 141; DB 2; Length 1388;  
Best Local Similarity 16.8%; Pred. No. 0.0002;  
Matches 94; Conservative 97; Mismatches 176; Indels 192; Gaps 20;  
QY 54 KDKTAISENLSFSEVGEIDAEVKKALGTGIKMKI-----MMERKEKEHTNLS 104  
DB 780 KOKOVLEEDVRNLT-----LKTEQFQKRLQNDLQKMQQVNTLKMSEKQLKQNNHLM 835  
QY 105 TLK-----KCREKEQALKLNEVQHLLEEE-----RLCRSLADSWGEC----- 145  
DB 836 ENKMLEKQNAELRKERQADQOMKELQDLQAEQYFSTLYKTQVRELKECEKTKLKG 895  
QY 146 ---RSCLENNCHRYTTQPSWSSVKNKIERFFRIYQFLPFPHEDNE--KDLPISE--- 197  
DB 896 ELQQRKQELQDRDSLAQALETTLTKADSEQLARSIAEQYSDLEKEKIMKELEIKEMA 955  
QY 198 ----KLIEDAOLQMEDVFSOLTVDVNSLFRS---FNVFRMQOEFQDTFQSHFISDT 250  
DB 956 RHKQELTEKDATIASLEETNRLTSDVANLANEKEELNKLKQVQQLSR-----LKDE 1009  
QY 251 DUTEYFFFPAPFKEPMTKADLQCCWDIPNFOLFNCFSVSYESVETITMKLKAIE--- 307  
DB 1010 EISAAAIKAQFEKOLLITERTLK-----TOAVNKLAEITMRKE 1046  
QY 308 -----DLPKQDAPDHGGLSKMLPGODRGLGCEL-----DONLSRCFFHEKCKQCO 355  
DB 1047 PVKRGNDTDVRREK-----ENRKLHMLKSEREKLTOQMIKTQKELNEMQ 1092  
QY 356 AHLSEDCPDVPALHTELDIAIRLVNSNQYQGIQLQMTKKHLE-----DTAY 402  
DB 1093 AQIAEE---SQIRIELQMTLDSKDSIDIELRSQALHIGLSSIGSGPGDAEADGGF 1148  
QY 403 LVEKMRG-----QFGWSEL-----A 418  
DB 1149 PESRLGWLSPVRNNTKFGWVKYVYVSSKKILFYDSEQDKQSNPYMWLDIDKLPHV 1208  
QY 419 NQAPETELIFNSIQVVPRI-----HEGNISKQDETMTDLSILSSNFTLKLPLEESAE 472  
DB 1209 RPVTQDVTDAKAKEIPRIQIILYANEGESKREQE-----FPPEPVGE 1251  
QY 473 SSNFI---GYVVAKALQHF 488  
DB 1252 KSNYICKHGHEFTPLYHF 1270

RESULT 13  
US-08-328-254-6  
; Sequence 6, Application US/08328254  
; Patent No. 5710022  
; GENERAL INFORMATION:  
; APPLICANT: Zhu, Xueliang  
; APPLICANT: Lee, Wen-Hwa

;; TITLE OF INVENTION: A NO. 5710022el Nuclear Mitotic Phosphoprotein  
;; NUMBER OF SEQUENCES: 8  
;; CORRESPONDENCE ADDRESS:  
;; ADDRESSEE: Campbell and Flores  
;; STREET: 4370 La Jolla Village Drive, Suite 700  
;; CITY: San Diego  
;; STATE: California  
;; COUNTRY: USA  
;; ZIP: 92122  
;; COMPUTER READABLE FORM:  
;; MEDIUM TYPE: Floppy disk  
;; COMPUTER: IBM PC compatible  
;; OPERATING SYSTEM: PC-DOS/MS-DOS  
;; SOFTWARE: PatentIn Release #1.0, Version #1.25  
;; CURRENT APPLICATION DATA:  
;; APPLICATION NUMBER: US/08/328,254  
;; FILING DATE: 24-OCT-1994  
;; CLASSIFICATION: 435  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/141,239  
;; FILING DATE: 22-OCT-1993  
;; ATTORNEY/AGENT INFORMATION:  
;; NAME: Campbell, Cathryn A.  
;; REGISTRATION NUMBER: 31,815  
;; REFERENCE/DOCKET NUMBER: P-CJ 1191  
;; TELECOMMUNICATION INFORMATION:  
;; TELEPHONE: (619) 535-9001  
;; TELEFAX: (619) 535-8949  
;; INFORMATION FOR SEQ ID NO: 6:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 2482 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-328-254-6

Query Match 5.0%; Score 130.5; DB 1; Length 2482;  
Best Local Similarity 18.8%; Pred. No. 0.0049;  
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;  
QY 45 KDSHCAPTWKDKTAISENLK-----SFSEVGE-----IDAE---EVKKALTGIQ 87  
DB 1502 KOLHAEKLLKRERENDSKDKVENLERELQMEENOELVILDAENSKAEVETLKTQIEE 1561  
QY 88 M-----KIMMERKEKH-----TNLMSTLKKCREKEQALKLLNEV 123  
DB 1562 MARSILKVFELDLVTLRSEKENITQIQEKQGLSELDKLLSFKSLLEEKQA-----EI 1616  
QY 124 QHLEEEERLCRESIADSWGECRSCLENNC--MRIYTTCPQSWs-----SVKNKIERF 174  
DB 1617 Q--IKEESKTAVEMQLNQKLNEVAALCGDQEIWKATEQSLDPPPIEEHQLRNSIEKL 1674  
QY 175 FRKIYQFLPFPHEDNEKDLPISEKLEED----- 203  
DB 1675 RARL-----EADKKQLVQLQKLESEHDLKGVLENLERLEIARTNQEHAALEA 1727  
QY 204 -----AOLQMEDVFSOLTVDVNSLFRSFNFRMQOEFQDTFQSHFISDTDL 252  
DB 1728 ENSKGEVETLAKITGMTQSLRGLDVTIRSEKENITNLOKEQER-----ISELEI 1781  
QY 253 TEPYFFPAPFKEPMTKADLQCCWDIPNFOLFNCFSVSYESVETITMKLKA--IEDLPK 311  
DB 1782 INSSPENILQKEQEKVQWKE-----KSTAMEMLQTLQKELNE 1820  
QY 312 QDKAPDHGGLSKMLPGODRGLGCELQNLs---RCFHEKCKQCOAHLSEDCPDVPAL 368  
DB 1821 RYVALHN-----DQEACKAKEQNLSSQVECLEL-EKAQLLQG----- 1856  
QY 369 HTDELDAIRLVNSNQYQGIQLQMTKKHLEDTAYLVEKMRGQFQWVSELANOAPETETIF 428  
DB 1857 ---LDEAKNNYIVLOSSVNGLIQ-----EVEDGKQLEKKDEE---ISRLKNIQIQDEQIV 1906

```
QY 429 NSIOVPRH---EGNISQDETMWD--LSILPSSNFTKIPLE 468
Db 1907 SKLSQVEGEHLWKEQNLEURLNLTVELEQIKVQLQSNASLODPTLE 1952

RESULT 14
US-08-533-306A-6
; Sequence 6, Application US/08533306A
; Patent No. 5837457
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/533,306A
; FILING DATE: September 25, 1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869COB
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-533-306A-6

Query Match 4.9%; Score 129; DB 2; Length 816;
Best Local Similarity 20.0%; Pred. No. 0.0012;
Matches 101; Conservative 76; Mismatches 175; Indels 154; Gaps 21;

QY 4 KAEKNGPSRSWQLHWGDIANNNGN--MKPPLLVIIVCLLWLDKSHCAPTWKDKTAISE 61
Db 73 QGEQQTSPR-----EYVDLEREAGVKYLPKAPMILGVCVIW-----KGWID----- 114
QY 62 NLKSFSEVGEIDADEE-----VKKALTGIKQMKIMMERKEKHTNLM-----STLKK 108
Db 115 -LQRLDGMCGLEFDEERAQOEDALAAQAFEAARRRTREFEDRDRSHREEMENEVESVTGM 173
QY 109 CREEKQOEALKL-----LNEVQEHLEEEERLC-----RESLADSWGEC 145
Db 174 LNEAEKAIKLAQDVASLSQQLDTQELLOETROKLVNSTKLRLEERNLSLQOLDEE 233
QY 146 RSLCLENCHRIYTTQPSWSSVKNKIERF-----FRKIYQFLFPFHEDNEKD 192
Db 234 MEA-KONLERHISTINQLDSKKLQDFASTVEALEEGKKRFQXEIENL---TQYEEK 289
QY 193 LPISKEKLIEDAQLTQMEDVFSQLTVDVNSLNFNSFNFRMQOQFDO-----TFQSHF 246
Db 290 AAAYDKLEKTKNRLOQELD---DLVVVDNQRQLVSNL-EKKQRKFDQLLAEKNISSKY 345

QY 247 ISDTDLTEPYFFAFSPKPMTKADLEQCQMDIPNFFQFCNFSVSIYESVSETITKMLKAI 306
Db 346 ADERDRAE-----AEAREKETKA-----LSLARALEEALAK 377
QY 307 EDLPKQDKAPDHGGLISKMLPGQDRGLCGELDONLSRCFKFHEKCKQAHLSDCPDVP 366
Db 378 EELER-----TNKMLKAEMEDLVSSKDDVGKNVHELEKSKRALETQMEEMKTQLE 427
QY 367 ALHTEL----DEAIRL-VVNSNQYQGILOMTKRLHLEDATYLVKMRGQFGWVSELANQA 421
Db 428 ELEDELQASEDAKLRLEVN-----MQALKGQF-----ERDLQA 460
QY 422 PETEIIFNSIQVVPRIHEGNISKQDE 447
Db 461 RDEQNEEKRRQLQRLHVEYTELEDE 486

RESULT 15
US-08-742-923A-6
; Sequence 6, Application US/08742923A
; Patent No. 5869611
; GENERAL INFORMATION:
; APPLICANT: Liu, Pu
; APPLICANT: Collins, Francis S.
; APPLICANT: Siciliano, Michael J.
; APPLICANT: Claxton, David
; TITLE OF INVENTION: Markers for Detection of Chromosome 16
; TITLE OF INVENTION: Rearrangements
; NUMBER OF SEQUENCES: 14
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Harness, Dickey & Pierce, P.L.C.
; STREET: P.O. Box 828
; CITY: Bloomfield Hills
; STATE: MI
; COUNTRY: USA
; ZIP: 48303
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/742,923A
; FILING DATE: No. 5869611member 1, 1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Smith, Deann F.
; REGISTRATION NUMBER: 36683
; REFERENCE/DOCKET NUMBER: 2115-00869DVC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (810) 641-1600
; TELEFAX: (810) 641-0270
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 816 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-742-923A-6

Query Match 4.9%; Score 129; DB 2; Length 816;
Best Local Similarity 20.0%; Pred. No. 0.0012;
Matches 101; Conservative 76; Mismatches 175; Indels 154; Gaps 21;

QY 4 KAEKNGPSRSWQLHWGDIANNNGN--MKPPLLVIIVCLLWLDKSHCAPTWKDKTAISE 61
Db 73 QGEQQTSPR-----EYVDLEREAGVKYLPKAPMILGVCVIW-----KGWID----- 114
QY 62 NLKSFSEVGEIDADEE-----VKKALTGIKQMKIMMERKEKHTNLM-----STLKK 108
Db 115 -LQRLDGMCGLEFDEERAQOEDALAAQAFEAARRRTREFEDRDRSHREEMENEVESVTGM 173
```

Qy 109 CREEKQALKL-----LNEVQEHLEEEERLC-----RESLADSWGEC 145  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 174 LNEAGRAIKLAKDVASISSQLQDTQELLQETRQKLNSTKLRLQLEERNLSLQDLDEE 233  
Qy 146 RSCLENCMRIYTCQPSWSVKNKIERF-----FRKIYQFLPPFHEDNEKD 192  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 234 MEA-KONLERHISTINIQLSDSKKKLQDFASTVEALEEGKKRFQKEIENL---TQYEEK 289  
Qy 193 LPISEKLEEDAQLTOMEDVFSOLTVDVNSLFRSENVFRQMOQEFDO-----TFQSHF 246  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 290 AAAYDKLEKTNRLQQLD---DLVVDLQORQLVSNL-EKKQRKFDOLLAEKNISKY 345  
Qy 247 ISDTDLTEPYFPFAPFSKEPMTKADLEQCWDIPNFFQLFCNFSVSIYESVSETITKMLKAI 306  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 346 ADERDRAE-----AEAREKETKA-----LSLARALEEALAK 377  
Qy 307 EDLPQDKAPDHGLISKMLPGDRGLCGELDONLSRCFHEKCKQCOAHLSEDCPDVP 366  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 378 BELER-----TNKMLKAEMEDLVSSKDDYGVKNVHELEKSKRALETQMEEMKTOLE 427  
Qy 367 ALHTEL-----DEAIRL-VNVSNOQYQILQWTRKHLEDYAVLVEKMRGQFGWSELANQA 421  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 428 ELEDELQASEDAKLRLEVN-----MQALKQGF-----ERDLQA 460  
Qy 422 PETEIIFNSIQVPRIHEGNISKODE 447  
| : : : : : | : : : : : | : : : : : | : : : : : |  
Db 461 RDEQNEKKRROLQOLHEVETELEDE 486

Search completed: July 2, 2002, 11:43:54  
Job time: 628 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:40:16 ; Search time 32.87 Seconds  
(without alignments)  
1447.040 Million cell updates/sec

Title: US-09-722-544a-2  
Perfect score: 2632  
Sequence: 1 MKIKAEKNEGPRSNWQLHW.....FTGYVAKALQHFKEHFKTW 495  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	16.5	449	1 A41386	clusterin precursor
2	416	15.8	451	2 I50131	clusterin - quail
3	410.5	15.6	445	2 A40018	clusterin precursor
4	406	15.4	449	2 S07714	T64 protein precursor
5	403	15.3	448	2 A40714	clusterin precursor
6	402	15.3	448	2 I56335	clusterin precursor
7	400.5	15.2	439	2 A35744	apolipoprotein J -
8	400.5	15.2	446	2 A42108	clusterin precursor
9	380	14.4	447	2 A27205	clusterin precursor
10	168	6.4	191	2 I48174	sulfated glycoprot
11	160	6.1	1738	2 T14867	interaptin - slime
12	151.5	5.8	1091	2 T34107	hypothetical prote
13	151	5.7	1526	2 T41522	myosin ii - fissio
14	146	5.5	1136	2 B70356	chromosome assembl
15	145.5	5.5	534	1 A44073	Cik1 protein - yea
16	145	5.5	1388	2 S70633	serine/threonine-s
17	144	5.5	3225	2 I52300	giantin - human
18	144	5.5	3259	1 A56539	giantin - human
19	142.5	5.4	2653	1 S28261	centromere protein
20	141.5	5.4	3433	1 S28381	utrophin - human
21	141	5.4	1005	2 A64465	hypothetical prote
22	140.5	5.3	2166	2 G70163	hypothetical prote
23	139	5.3	1955	2 T30934	myosin-like prote
24	137	5.2	1084	2 G71329	hypothetical prote
25	137	5.2	1300	2 I53799	Cg1 protein - huma
26	137	5.2	1940	1 S04090	myosin heavy chain
27	136	5.2	1132	2 T00259	hypothetical prote
28	135	5.1	1060	1 A40264	kinesin-related pr
29	134.5	5.1	1957	2 T38077	hypothetical colle

## ALIGNMENTS

## RESULT 1

A41386

clusterin precursor [validated] - human  
N:Alternate names: apolipoprotein J; Complement cytotoxic inhibitor SP-40; complement TRPM-2/clusterin protein  
N:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Homo sapiens (man)  
C:Date: 03-Apr-1992 #sequence\_revision 17-Nov-1995 #text\_change 08-Dec-2000  
C:Accession: S43646; S04662; A41386; A35833; S34056; A53177; A37816; B37816; PL0136; R:Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M.  
Eur. J. Biochem. 221, 917-925, 1994  
A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with A:Reference number: S43646; MUID:94237156  
A:Accession: S43646  
A:Molecule type: DNA  
A:Residues: 1-449 <NON>  
A:Cross-references: GB:M64722; NID:g339972; PIDN:AAB06508.1; PID:g339973 R:Kirszbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.; EMBO J. 8, 711-718, 1989  
A:Title: Molecular cloning and characterization of the novel, human complement-associ A:Reference number: S04662; MUID:89251601  
A:Accession: S04662  
A:Molecule type: mRNA  
A:Residues: 1-449 <KIR>  
A:Cross-references: EMBL:X14723; NID:g30250; PIDN:CAA32847.1; PID:g30251  
A:Note: parts of this sequence, including the amino end of the mature protein, were c R:Jenne, D.E.; Tschopp, J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989  
A:Title: Molecular structure and functional characterization of a human complement cy tis fluid.  
A:Reference number: A41386; MUID:89386692  
A:Accession: A41386  
A:Molecule type: mRNA  
A:Residues: 2-449 <JEN>  
A:Cross-references: GB:M25915; NID:g180619; PIDN:AAA35692.1; PID:g180620 R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J.  
Biochemistry 29, 5380-5389, 1990  
A:Title: Apolipoprotein J: structure and tissue distribution.  
A:Reference number: A35833; MUID:90344779  
A:Accession: A35833  
A:Molecule type: mRNA  
A:Residues: 34-449 <DES>  
A:Cross-references: GB:J02908; NID:g178854; PIDN:AAA51765.1; PID:g178855 R:Ghiso, J.; Matsubara, E.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T. Biochem. J. 293, 27-30, 1993  
A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's A:Reference number: S34056; MUID:93319521  
A:Accession: S34056  
A:Molecule type: protein  
A:Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI> R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse Arterioscler. Thromb. 11, 645-652, 1991



## C:Genetics:

A:Gene: T64

A:Introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2

C:Superfamily: clusterin

Query Match 15.8%; Score 416; DB 2; Length 451;

Best Local Similarity 25.7%; Pred. No. 1.2e-18;

Matches 118; Conservative 93; Mismatches 180; Indels 68; Gaps 15;

Qy 63 LKSFSEVEIDADDEVKALTGTQKIMMERKEHTNLMSTIKKCRKQKALKLLNE 122

Db 27 LKQLSAAGSKYIDAENAINGVKMTLMDKTSKEHQAMLHTEETKKKEEAVKLALE 86

Qy 123 VOEHLEERLCRESLADSGECSLENNCMRIYT--TCQPSWSSVKNKITERFRKIYQF 181

Db 87 KEKQAEKQEVNCTMLSLWEECKPCLKHTCMRVYSKMGSGGLVGRQLEEFNRSPPF 146

Qy 182 LFPFHEDNEKDLPISEKLEIE-----DAQLTQMEDVFSQLTVDVNSLNFNSVFMOMQ 236

Db 147 SIWNGERIDDLREORQERRFDELRGLMEDGVEDIFQDSTQLYGPAFFFR----- 202

Qy 237 EFDQTFQSHFISDLDTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

Db 203 -----TPPFGGPREAF--VPPVQRVHL-----VPRRLSRELHPFFQHPM 240

Qy 286 NFSVSIYESVSETTKMLKAIEDLPKQKADPH--GGLISKMLP--GQDRGLGELDONLS 342

Db 241 HGFHRLFOPLFQMTQHMLDG-----HGAWHEPLGGFATESRNFSTDRMVCREIRNSA 294

Qy 343 RCFKHFECORCOAHLSDC----PDVPALHTELDEAIRLVNSNOOYGQILQMKRHL 398

Db 295 GCLRMRDCEKREILAVDCQTPVQSOLREQFEDALRLAERTTRYDILLSAFQAEML 354

Qy 399 DTAYLVEKMRQGFQGVSELANQAP-----ETEIFNSIQVVPRIHEGNISKQDETMMTD 452

Db 355 NTSSLLDQLNRQFGVSRGLNTQGDNGFLQVTTVFESK---TPNLEDP--SAPADTQVT- 408

Qy 453 LSLIPSSNFTLKIPLEESAESSNFTGYVAKALQHFKEH 491

Db 409 VOLFDSEPLSLTVPGDISMDPRFMEIVAQALQHYKQN 447

## RESULT 3

A40018

Clusterin precursor - dog

N:Alternate names: 80k secreted glycoprotein, renal; complement cytotoxic inhibitor

C:Species: Canis lupus familiaris (dog)

C:Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 13-Aug-1999

C:Accession: A40018

R:Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pilarsky, C.; Appel, D.; Ha

J. Biol. Chem. 266, 9924-9931, 1991

A:Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in

F:1-445/Domain: clusterin alpha chain #status experimental &lt;ACH&gt;

F:86,103,145,277,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

A:Accession: A40018

A:Molecule type: mRNA

A:Residues: 1-445 &lt;HAR&gt;

A:CROSS-references: GB:M55251; GB:M38757; NID:g163953; PIDN:AAA30846.1; PID:g163954

A:Note: parts of this sequence, including the amino ends of the alpha and beta chains,

C:Superfamily: clusterin

F:1-32/Domain: signal sequence #status predicted &lt;SIG&gt;

F:23-226/Domain: clusterin beta chain #status experimental &lt;BCH&gt;

F:23-226,227-445/Product: clusterin #status experimental &lt;MAT&gt;

F:227-445/Domain: clusterin alpha chain #status experimental &lt;ACH&gt;

F:86,103,145,277,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match

Best Local Similarity 23.9%; Pred. No. 2.7e-18;

Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

Qy 30 MKPPLLVFTVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDAD-----EEVKKALT 83

Db 1 MKKTLILLVGLLL-----TWNGRVLGQAVSDTELOEMSTEGSKYINKKLNALK 51

Qy 84 GIKQKIMMERKEHEHTNLMSTIKKCRKQKALKLLNEVQEHLEERLCRESLADSG 143

Db 52 GVKQIKTLIEQTNEERKSLLSLEAKKKEDALNDTKDSETKLKASQGVCDTMMALWE 111

Qy 144 ECRSCLENNCMIYT--TCQPSWSSVKNKITERFRKIYQFLFPFHEDNEKDLPISEKLEIE 202

Db 112 ECKPCLKQCMKPYARVCSGLVGHQLEEFNLQSSPYFWNGDR-----IDSLLEN 165

Qy 203 DAQLTQMEDVFSQLTVDVNSLNFNSVFMOMQOEFDQTFQSHFIS-----DIDLTEPYFF 258

Db 166 DRQOATHADV-----MQDSFNRSI-----MDLFFQDRFFTRPDQTYHSP--F 209

Qy 259 PAFSKEPMTKADLEQCWDI-----PNFFQFCNFSVSIYESVSETTKMLKALE--- 307

Db 210 SLFORPFPNPKFRIARNIIPFRFQPLNFHDMFQFFEDMIHOA--QOAMDVNLHRIYHF 268

Qy 308 --DLPKODRAPDHGGLISKMLPGQDRGLGELDONLSRCFKFHEKQKQCAHLSEDC--- 362

Db 269 PIEFPED-----NRTVCKEIRHNSGTGLKMKDOCEKCOEILSVDCSSN 312

Qy 363 -PDVPALHTELDEAIRLVNSNOOYGQILQMKRHLDEATYLVKMRQGFQGVSELANQA 421

Db 313 NPAQVOLRQELSLSQIAEKFTKLYDELQSYQEKMFNTSSLLKQLNEQFSWYSQANLT 372

Qy 422 PETEIFNSIQVVPRIHEGNISKQDETMMTDLSILPSSNFTLKIPLEESAESSNFTGYV 480

Db 373 QSDPPYLVQTTVGSQTSDSNPVPGTKVV--VKLFSDSDPITVMIPEAVSRNNPKFMTV 430

Qy 481 VAKALQHFKE 490

Db 431 AERKALQERYQ 440

## RESULT 4

S07714

T64 protein precursor - Japanese quail

C:Species: Coturnix coturnix japonica (Japanese quail)

C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 13-Aug-1999

C:Accession: S07714

R:Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.

A:Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b

Oncogene Res. 4, 127-136, 1989

A:Reference number: S07714; MUID:89239492

A:Accession: S07714

A:Molecule type: mRNA

A:Residues: 1-449 &lt;MIC&gt;

A:CROSS-references: EMBL:X15825; NID:g62594; PIDN:CAA33823.1; PID:g62595

C:Superfamily: clusterin

F:1-18/Domain: signal sequence #status predicted &lt;SIG&gt;

F:19-449/Product: T64 protein #status predicted &lt;MAT&gt;

Query Match

Best Local Similarity 25.7%; Pred. No. 5.1e-18;

Matches 117; Conservative 91; Mismatches 180; Indels 68; Gaps 15;

Qy 63 LKSFSEVEIDADDEVKALTGTQKIMMERKEHTNLMSTIKKCRKQKALKLLNE 122

Db 27 LKQLSAAGSKYIDAENAINGVKMTLMDKTSKEHQAMLHTEETKKKEEAVKLALE 86

Qy 123 VOEHLEERLCRESLADSGECSLENNCMRIYT--TCQPSWSSVKNKITERFRKIYQF 181

Db 87 KEKQAEKQEVNCTMLSLWEECKPCLKHTCMRVYSKMGSGGLVGRQLEEFNRSPPF 146

Qy 182 LFPFHEDNEKDLPISEKLEIE-----DAQLTQMEDVFSQLTVDVNSLNFNSVFMOMQ 236

Db 147 SIWNGERIDDLREORQERRFDELRGLMEDGVEDIFQDSTQLYGPAFFFR----- 202

Qy 237 EFDQTFQSHFISDLDTEPY--FFPAFSKEPMTKADLEQCWDIPN-----FFQLFC 285

```
Db 203 -----TPPGGFREAF-VPPVQVRVHL-----VPRRLSRELHPFFOHPM 240
Qy 286 NFSVSIYESVETTKMLKAIEDLPKODKAPDH--GGGLSKMLP-GQDRGLGELQNLNS 342
Db 241 HGFRHLFQPLFEMTQHMLDG-----CHGAWHPLGGGFATESRNFSTRMVMCREIRNSA 294
Qy 343 RCRFHEKCKCQQAHLSEDC-----PDVPALHTDELDEAIRLVNVSNOQYGOILOMTRKHLE 398
Db 295 GCLMRDECBKREILAVDCSQTDPVQSQRREQFEDALRAERFTRRYDILLSAFAQAEML 354
Qy 399 DTAYLVKMRGQFGWSELANQAP-----ETELIFNSIQVPRRIHEGNISKODETMMTD 452
Db 355 NTSSLLDLNRFQGWSELNLTQNGDGFLOVTVFSK-----TPNLEDP--SAPADTQVT- 408
Qy 453 LSILPSSNFTLKIPLESAESSNFYGVVAKALQHF 488
Db 409 VOLFDSEPLSLTPGDISWDPRFMEITVPEQALQHY 444

RESULT 5
A40714
clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
A:Accession: A40714; JN0699
R:French, L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <PRE>
A:Cross-references: GB:L08235; NID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LEE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <Sig>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-448/Domain: clusterin alpha chain #status predicted <MAT>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.3%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 7.8e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 34 LLVFIIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGIKQ 87
Db 3 ILLICVALLI-----WDNGMVLGQEVSDNELQELSTQGSRYINKEIQNAVQGVKH 54
Qy 88 KIMMERKEKEHTNLMSTLKCKREBKQKALKLLNEVQEHLEEEERLCRESLADSGEGRS 147
Db 55 IKTLIEKTNARSKSLNLEEAKKKKEDALETRDSEMKLKAFFPVCNETMMALWEECKP 114
Qy 148 CLNNCMRIYT--TCQPSWSSVKNKTERFRKIYQFLFFPHEDNEKDLIPSEKLEEDAQ 206
Db 115 CLKHTCMFYARVCRSGSLVQGLVQQLLEEFNLNOSPFFYFWMNGDR-----IDSLLESDRQ 168
Qy 207 TQ-----MEDVFSQLTVDVNSLFRNFVFRMQQEFDTQFSHFISDITLTPYFFP--- 259
Db 169 SQVLAMDQDSFARAGIIDTLFDQDF--FAR-----ELHDPHYFS-----PIGFPHKR 214
```

```
Qy 260 ---AFSKEPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVSETITTKMLKAIEDLPKQDKA 315
Db 215 PHFLYPKSLRSLRSLMSPSHYGPSPFNMMQPFPEMLHQAAQMDVOLHSPAQFPDQD--- 272
Qy 316 PDHGLISKMLPGQ-DRGLCGELDQNLSCFKFHEKQCKQAHLSDEC-----PDVPALHT 370
Db 273 -----FLREGEDDRTVCKEIRRNSTGCLKMKGCKQCEILSVDCSTNNPAQANLRQ 324
Qy 371 ELDEAIRLVNVSNOQYGOILOMTRKHLEDTAYLVKMRGQFGWSELANQAPETETIIFNS 430
Db 325 ELNDSLQVAERTEQYKELLQSFQSKMLNTSSLLQNDQFNWVSQNLANTQGEDKYILR 384
Qy 431 IQVPRRIHEGNISKODETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQHF 489
Db 385 VSTV-TTHSSDSEVPSSRVTEVVVKLFDSPTITVLPPEVSKDNPKFMDTVAEKALQEYR 442

RESULT 6
156335
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
A:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Aronow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A>Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
A:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.3%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 9.1e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 30 MKPPLLVTIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKALTGIKOM 88
Db 1 KMLLVCVAMLLIW--DNGMV---LGEQVSDNELQELSTQGSRYINKEIQNAVQGVKHI 55
Qy 89 KIMMERKEKEHTNLMSTLKCKREBKQKALKLLNEVQEHLEEEERLCRESLADSGEGRS 148
Db 56 KTLIEKTNARSKSLNLEEAKKKKEDALETRDSEMKLKAFFPVCNETMMALWEECKPC 115
Qy 149 LENNCMRIYT--TCQPSWSSVKNKTERFRKIYQFLFFPHEDNEKDLIPSEKLEEDAQ 207
Db 116 LKHTCMKFYARVCRSGSLVQGLVQQLLEEFNLNOSPFFYFWMNGDR-----IDSLLESDRQ 169
Qy 208 Q-----MEDVFSQLTVDVNSLFRNFVFRMQQEFDTQFSHFISDITLTPYFFP--- 259
Db 170 QVLAMDQDSFARAGIIDTLFDQDF--FAR-----ELHDPHYFS-----PIGFPHKR 215
Qy 260 ---AFSKEPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVSETITTKMLKAIEDLPKQDKAP 316
Db 216 HFLYPKSLRSLRSLMSPSHYGPSPFNMMQPFPEMLHQAAQMDVOLHSPAQFPDQD--- 272
Qy 317 DHGGLISKMLPGQ-DRGLCGELDQNLSCFKFHEKQCKQAHLSDEC-----PDVPALHT 371
Db 273 -----FLREGEDDRTVCKEIRRNSTGCLKMKGCKQCEILSVDCSTNNPAQANLRQ 325
Qy 372 LDEAIRLVNVSNOQYGOILOMTRKHLEDTAYLVKMRGQFGWSELANQAPETETIIFNSI 431
Db 326 LNDLSLQVAERTEQYKELLQSFQSKMLNTSSLLQNDQFNWVSQNLANTQGEDKYILRV 385
Qy 432 QVPRRIHEGNISKODETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQHF 489
Db 386 STV-TTHSSDSEVPSSRVTEVVVKLFDSPTITVLPPEVSKDNPKFMDTVAEKALQEYR 442
```



```
RESULT 7
A35744
clusterin precursor - bovine
N:Alternate names: complement cytotoxicity inhibitor; glycoprotein III
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Bos primigenius taurus (cattle)
C:Date: 05-Oct-1990 #sequence_revision 05-Oct-1990 #text_change 13-Aug-1999
C:Accession: A35744
R:Palmer, D.J.; Christie, D.L.
J. Biol. Chem. 265, 6617-6623, 1990
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin cells
A:Reference number: A35744; MUID:90216681
A:Accession: A35744
A:Molecule type: mRNA
A:Residues: 1-439 <PAL>
A:Cross-references: GB:J05391; NID:g163114; PIDN:AAA30554.1; PID:g163115
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were determined from complementary DNA
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein
F:1-19/Domain: signal sequence #status predicted <SIG>
F:20-221/Domain: clusterin beta chain #status experimental <BCH>
F:20-221,222-439/Product: clusterin #status experimental <MAT>
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted
Query Match 15.2%; Score 400.5; DB 2; Length 439;
Best Local Similarity 24.5%; Pred. No. 1.1e-17;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;
QY 52 TWKDKTATSE-NLKSFSEGEIDAEVKKALGTGKIMMERKEKEHTLMSTLKKCR 110
DB 13 SWESGALSDELQEMSTEGSKYVKNKALKEVKQIKTQIEQNEERKLLLSLEAK 72
QY 111 EEQBALKLLNEVQPHLEEEERCLRESLADSWGECRSCLNCKMIYT-TCQPSWSSVKN 169
DB 73 KKKEDALNDTRDSNENKLRASQVCNETMTALWEECKPCLKOTCMKFYARVCRSGSLGVGH 132
QY 170 KTERFERKIYQFLPFPHEDNEKDLISEKLIIEEDALQTMEDVFSQTLVDVNSLFRSN 229
DB 133 QLEEFLNSSPFYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSTRASS 179
QY 230 VFRMQOEFQDTQFSHFIF-----SDTDLTEPY-----FF---PAFSKEPMKADLQEC 274
DB 180 I-----MDELFDQDFLRRLRQDTQYYSFSSFPGRSILFFNPKSRFARNVMPFLLEPF 232
QY 275 WPIPNFQLFNFSVSIYESVSETITKMLKAIETDLPKQKAPDHGGLSKMLPGQDRGLC 334
DB 233 ----NFHDVFQPF----YDHIHQAOQAMDAHLQ-----RTPYHPPTM-EFTENNDRTVC 277
QY 335 GELDONLSRCFHEKCKCOAHLSEDC----PDVPAHLTEDEAIRLVNVSNOOYGOIL 390
DB 278 KEIRHNSGCLRMKQDCEKCEILEVDCSASNPOTTLRLQNLASLIQAEKFSRLYDQLL 337
QY 391 QMTRKHELTAYLVKMRQFGWVSLANOAPETE-----IFNSIQVVPRIHEGNTSK 444
DB 338 QSVQOKMLNTSALLKQLNQFTVWSOLANLTQSDQHYLVFTVNSHNSDPSIPSG---- 393
QY 445 QDETMTDLSILPSSNFTKIPLESAESSNFIGYVVAKALQHF 489
DB 394 -----LTKVIVLFNSFPITVTPQREVSPNFMENVAEKALQOYR 433
RESULT 8
A42108
clusterin precursor - pig
N:Alternate names: complement cytotoxicity inhibitor; Cp40 protein
C:Species: Sus scrofa domestica (domestic pig)
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 13-Aug-1999
C:Accession: A42108; JC5535; PC4475
R:Dlemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.
J. Biol. Chem. 267, 5257-5264, 1992
A:Title: Expression of porcine complement cytotoxicity inhibitor mRNA in cultured aortic
A:Reference number: A42108; MUID:92184774
A:Accession: A42108
A:Molecule type: mRNA
A:Residues: 1-446 <DIE>
A:Cross-references: GB:M84639; NID:g164408; PIDN:AAA31013.1; PID:g164409
A:Experimental source: aortic smooth muscle cells
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.
Biochem. Biophys. Res. Commun. 234, 712-718, 1997
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary
A:Reference number: JC5535; MUID:97318844
A:Accession: JC5535
A:Molecule type: DNA
A:Residues: 1-446 <OGA>
A:Accession: PC4475
A:Molecule type: protein
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>
A:Experimental source: pituitary gland
C:Superfamily: clusterin
F:1-22/Domain: signal sequence #status predicted <SIG>
F:23-227/Domain: clusterin beta chain #status predicted <BCH>
F:23-227,228-446/Product: clusterin #status predicted <MAT>
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>
Query Match 15.2%; Score 400.5; DB 2; Length 446;
Best Local Similarity 24.4%; Pred. No. 1.1e-17;
Matches 116; Conservative 98; Mismatches 196; Indels 71; Gaps 16;
QY 34 LLVFIIVCLLWKDSHCAPTWK-----DKTAISENLKSFSEGEIDAEVKKALGTGK 86
DB 4 LLLLVGLL-----TWNGPWVLGDKRAISKELQENSTEGSKYVKNKIKALAEVK 54
QY 87 QMKIMMERKEHTNLMSTLKKCRKQKALKNVEQHELEERCLRESLADSWGECR 146
DB 55 QIKTLIEQSNSEERKLLSLEEAARKKEDALNDTRDTETKLGSGGLCNFTMALWEECK 114
QY 147 SCLNNCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIIEDAQ 205
DB 115 PCLAKOTCMKFYARVCRSGSLGVGHQLEFLNQSSPFYFWINGDR-----IDSLMENDRQ 168
QY 206 LTOMEDVFSQTLVDVNSLFRSNFVFRQOQEFQTFQSHFIS-----DTDLTEPY----- 256
DB 169 QSHVMDI-----MEDSFNRASNI-----MDELFDQDFRNFPEFDTQFSPFGSSHR 214
QY 257 --FPAPFSKEPMKADLEQCDWIPNPFQFCNFSVSIYESVSETITKMLKAIETDLPKQD 313
DB 215 GSLFNPKSRFARNIMPFLFTDL--NYHDMFQPF-----FDMHQAOQAMDAHLHRIYH- 268
QY 314 KAPDRGGLISKMLP--GQDRGLCGELDONLSRCFHEKCKCOAHLSEDC-----PDVPA 367
DB 269 -FPEAG-----VPNSNDRVAVCKEIRHNSGCLRMKQDCEKCEILSVDCSASNSOMQ 321
QY 368 LHTELDEAIRLVNVSNOOYGOILQMTKRKHELTAYLVKMRQFGWVSLANOAPETEI 427
DB 322 LRQELTYTSLQAEKFSKLYDQLLQSYQOKMLNTSLLKQLNEQFWSWSOLANLTQNDRY 381
QY 428 FNSIQVVPRIHEGNTSKQDETMTDLSILPSSNFTKIPLESAESSNFIGYVVAKALQHF 487
DB 382 YLQVTVV-NSHGSDPSVPSGLTKVVKLVFDSYPTILLIIPQEV--DPKFMETVAEALQ 438
QY 488 FKE 490
DB 439 YRQ 441
RESULT 9
A27205
clusterin precursor - rat
N:Alternate names: SGP-2; sulfated glycoprotein 2; TRPM-2
C:Species: Rattus norvegicus (Norway rat)
C:Date: 21-May-1988 #sequence_revision 21-May-1988 #text_change 13-Aug-1999
```

```
Query Match          6.4%; Score 168; DB 2; Length 191;
Best Local Similarity 35.1%; Pred. No. 0.0013;
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps

QY   330 DRGLGCLDQNLRSRCFKPFKEKCKQCAHLSDEC----FDVPALHTLDELDEATRLNVNSNQ 385
      || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    69 DRAVCKEIRHNSTGCULKMKGCKEQCEILSYDCSANNPAQAHLRQELNDLSLQVAERLTQR 128
      || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY   386 YGOILOMTKRKHLEDAYLVLEKMRGQFGWVSELAN 419
      || : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db    129 YNELLHSLOTKMLNTSSLLLEQLNEQFNWVSQLAN 162
```

RESULT 11

T14867

interaptin - slime mold (Dictyostellum discoideum)

C:Species: Dictyostellum discoideum

C:date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000

C:Accession: T14867

R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.

J. Cell Biol. 142, 735-750, 1998

A:Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily

A:Reference number: Z18248; MUID:983653468

A:Accession: T14867

A:Status: Preliminary; translated from GB/EMBL/DBDJ

A:Molecule type: DNA

A:Residues: 1-1738 <RIV>

A:Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC3458

C:Genetics:

A:Gene: abpD

A:Introns: 173/2; 1680/1

Query Match	Score 160;	DB 2;	Length 1738;
Best Local Similarity	6.1%;	20.9%;	
		Pred. No. 0.056;	

Qy	158	TTQCPWSVSVKNIERFFRKIIQVLPFPFHEDNEKDLPTSEKLIEDAQITQ-----	208
Db	401	ROGEDSARSEERATKEEIKKLKSQVLOQLEQDLQKKRVE---LTEQRKVLESKA	457
Qy	209	-MEDVRSQITVDVNSLNFNSFVFRQMOQEF-----DQTFQ-----	243
Db	458	SVADEFTGLMSSUNSL--REEN--ROYEEETRSLQTNIRTLODEVYQHODAITEKKNRAE	513
Qy	244	-----SHFIS-----DTLTEPYFPYPAFSKEPMTKADLEQWCIDPFPQFCNFS	288
Db	514	KAEYEIKENHRVQNASSSHDADIITLENEKTMQEEALEKAQEKDQALRE-----	568
Qy	289	VSIYE-----SVSETITKMLKAIEDLPKQKADPHGGLISKMI PGDRLGCGELDONLSRC	344
Db	569	VRVKKREMTASITSDRQIOSLAE--KVDS-----LTRELESSRRRR--EOLQEDQTKF	618
Qy	345	FKHEKCKQCAHLSDCQPDVPALHTELDEAIRLVNSNQYGOILQMTKHKLEDTAYILV	404
Db	619	LGSHD-----ETKAEWMKD-----LHEAQDEIEKLTNQAGOLKSNETLTLE--	668
Qy	405	EKMKGQFGW----SELANAQPETEITFIENSIQVVPRIHSGINSKODETMTMTDLSILPSSN	460
Db	669	ERLKAQAEKADKKYEETKVQLRAEDLADLQAA--QILSGNVESKFSMQKE-----	719
Qy	461	FTLKIPLEESAESSNFTGYVYVAKALQHFKEHFK	493
Db	720	--SKIERETILDNHN-----KELEKLRLEEK	743

RESULT 13  
T41522  
myosin ii - fission yeast (*Schizosaccharomyces pombe*)  
C:Species: *Schizosaccharomyces pombe*  
C:Date: 03-Dec-1999 #sequence revision 03-Dec-1999 #text change 08-Sep-2000

```
A>Status: preliminary; translated from GB/EMBL/DDBJ
A:Molecule type: DNA
A:Residues: 1-1526 <WOO>
A:Cross-references: EMBL:AL049498; PIDN:CAB39901.1; GSPDB:GN00068; SPDB:SPC
A:Experimental source: strain 972H; cosmid c645
C:Genetics:
A:Gene: SPDB:SPCC645.05c
A:Map position: 3
C:Superfamily: myosin heavy chain; myosin motor domain homology
F:80-743/Domain: myosin motor domain homology <MCO>

Query Match          5.7%   Score 151; DB 2; Length 1526;
Best Local Similarity 20.7%   Pred. No. 0.18;
Matches 93; Conservative 81; Mismatches 221; Indels 54; Gaps

QY    65 SFSEVGEIDADEVKKALTGIKQMKNMERKEEHTNMSTLKCKREKEQALKLLNEVQ 124
      |||| : | :: |:: : | : | : | : | : | : | : | : | : | : | : |
DB    895 SFSFETQ--QENIQRESASLKQINNELLESUKTSKVETL-----LSEQNEIUK 942
```

943	Db	EKLSLEEK----	DLDDTKGELESIRENNATVLSKAEFNEQCKSLQETIVTKDAELDCLKT	959
176	QY	RKVIQFLPFFPHEDNEKOLPISEKLIIEEDAOLTOMEDYFSQLTVDVNSLFNRSNFVRMQ	235	
999	Db	KYISDYKTEIQEMRLTNQMKNEKSIQEGSLSBSLKRKKLERE-NSTLLSDYSILKQOK	105	
236	QY	QEFD--QTFQSHFISDIDLTEPYFPFFAFSRKPMTKADLEQCWDIPNPFQFCNFSYSIYE	293	
1058	Db	EELSVLKGQVELTINNLEKYNYLEADVKQLPKLKELESINDKQDLYQLQATKNKELEA	111	
294	QY	SVSETITKMLKAIEDLP-KODKAPDGHGLISKMLPGQDRGICGELDQNL----	348	
		SRCFKFKH		

Db 1118 KVKECLNNIKSLTKELENKEKQNLSDASLKYYIELQ-----EIHENLLLVKVSLENYK 1171

Qy 349 EKQKQCOAHLSEDCPDVPAHTELDIAIRLVNVSQ-----QYQILOMTRKHLEDTA 401

Db 1172 KKEYGLQIDL-EGKQVDVTFNQELSKKHRLDTFNHESLLRSQASYKEKLSLASSENKDL 1230

Qy 402 YLVEKMRGOFQWSELANOAPETELIENSIOVVPRIHE-GNISQDFTMTDLSILPSSN 460

Db 1231 NKVSSLTKVQNELSPKASKVPELE-----RKITNLMEHYSQLGKTFDEKRRKALIASRDN 1285

Qy 461 FTLKIPLEESAESNFICYVAKALQHFK 489

Db 1286 BELR-SLKSELESKRKLEVEYQKVLVEVK 1313

RESULT 14

B70356

chromosome assembly protein homolog - Aquifex aeolicus

C;Species: Aquifex aeolicus

C;Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 02-Jun-2000

C;Accession: B70356

R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

Nature 392, 353-358, 1998

A;Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A;Reference number: A70300; MUID:98196666

A;Accession: B70356

A;Status: preliminary; nucleic acid sequence not shown; translation not shown

A;Molecule type: DNA

A;Residues: 1-1156 <AQF>

A;Cross-references: GB:AE000699; NID:g2983238; PIDN:AAC06839.1; PID:g2983243; GB:AE00069

A;Experimental source: strain VF5

C;Genetics:

A;Gene: xcpC

C;Superfamily: chromosome segregation protein SMC1

Query Match 5.5%; Score 146; DB 2; Length 1156;

Best Local Similarity 20.0%; Pred. No. 0.26;

Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;

Qy 63 LKSFSEVGETD-----ADEVKKALGIKQIMME-----RKEKHTNLMSTLKK 108

Db 163 IEEISGIGYEYERKKEALEELAEVELKIKEIDILIEISNOLKRLKEEKEKLEKELQ 222

Qy 109 CREKEQALKLNEVOHLEERLCRE--SLADSWGECRSCLENNCMRIYTTCPQSWSS 166

Db 223 IKRE-TEAKILLKEKELKERERIINELSSLESLEDTITFQIENEKEL----- 271

Qy 167 VKMKIERFFRKIYQFLFPFHEDNEK---DLPISEKLI-BEDAQLTQMEDVFSOLTVDVNS 222

Db 272 --NERERLLKVEKIMPFKEKVGKFTAETENAEISIKERELKESENVRKNLEELINN 329

Qy 223 LFNRSFNVRQ---MQOEFDTQFQSHFISDTLTPEY-FFPAFSKEPMTKADLEQCWDIP 278

Db 330 LUSDKENLEREVGTQLQLEK-----LKEEYKSLKEVEYERKLELEEEE----- 373

Qy 279 NFQQLCFNF-SVSIYGSVSETIKMLKATDLPKQDKAPDHGGLISKMLPGQDRGLCGEL 337

Db 374 --ERLAIITDEVYKLEEEKEKLEKLNLSL-NKEQKEQLTORANLNKKI-----ERI 421

Qy 338 DQNLSCFCFHEKQKCOQAHLSDCPDVPALHTELDIAIRLVNVSNOQYGOILOMTRKHL 397

Db 422 KEDINKLISEREKIKERKEQEIKRKAIAKKKEBEELNLTQELNIYEKRLSEVRKLL 481

Qy 398 E-----DTAVLVEKMRGOFQWVSELAN-QAPE----- 423

Db 482 EEVLKEKGAIERVRSFSDVSDFKDKIGYGSVELIRVKNPEHITATEVAGGRLKFI 541

Qy 424 ----TEIIFNSIQVVPRIHEGNIS 443

Db 542 VVEDEEVAKECICQAKRMNLGRFS 565

RESULT 15

A44073

CIK1 protein - yeast (Saccharomyces cerevisiae)

N;Alternate names: protein YM9646.11; protein YMR198w

C;Species: Saccharomyces cerevisiae

C;Date: 30-Sep-1993 #sequence\_revision 08-Mar-1996 #text\_change 21-Jul-2000

C;Accession: A44073; S50927

R;Page, B.D.; Snyder, M.

Genes Dev. 6, 1414-1429, 1992

A;Title: CIK1: a developmentally regulated spindle pole body-associated protein impor

A;Reference number: A44073; MUID:92354905

A;Accession: A44073

A;Molecule type: DNA

A;Residues: 1-594 <PAG>

A;Cross-references: EMBL:M96439; NID:g171221; PIDN:AAA34494.1; PID:g171222

A;Note: sequence extracted from NCBI backbone (NCBIN:110670, NCBIP:110672)

R;Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, January 1995

A;Reference number: S50917

A;Accession: S50927

A;Molecule type: DNA

A;Residues: 1-594 <PEA>

A;Cross-references: EMBL:Z47815; NID:g6422280; PIDN:CAA87820.1; PID:g6422291; GSPDB:GNO

C;Genetics:

A;Gene: SGD:CIK1; MIPS:YMR198w

A;Cross-references: SGD:S0004811; MIPS:YMR198w

A;Map position: 13R

C;Superfamily: CIK1 protein

C;Keywords: coiled coil

Query Match 5.5%; Score 145.5; DB 1; Length 594;

Best Local Similarity 20.9%; Pred. No. 0.12;

Matches 84; Conservative 71; Mismatches 138; Indels 109; Gaps 17;

Qy 44 LKDSHCAPTWKOKTAISENLKSFSEVG-----ETDADEEVKKA----- 81

Db 147 LKDNQCDLQKKNVLSKKNIVSHAVKQVEFENDLEELSNAKREWTYKLMVEVNLKPDE 206

Qy 82 -LTG-ITQMKIMME---RK-----EKEHTNLMSTLKKCKREEKQAL-KLLNEVOEHLE-EE 130

Db 207 RLTDENRQLKTEEEVNRKLFILQNEENENECKYKKELDKKFEIFKKVNDARIEDJGQ 266

Qy 131 ERLCR--ESLADSWGECRSCLENNCMRIYTTCPQSWSSYKNIERFFRKIYQFLFPFHED 188

Db 267 ERLSKVLKDLQDTHGELKENIK-----TC-----RDEFNDEKRIAGEAVNFHSM 311

Qy 189 NEKDLPISEKLTIEDAQLTQMEDVFSOLTVDVNSLNFNSFNVRQMQOEFDTQFSHFIS 248

Db 312 ELAVVPLKKLSTASTQALTVQVEEKQKQVEGANNNKKYVNELEKVVQEL----- 361

Qy 249 DTDLTPEYFPFAPSKEPMTKADLE-----QCWDIPNFQFCNFVS-----IYESVSET 298

Db 362 -----YTRQNLAISIEIKGYTRCFAYANERQMPDEHINLYVDRICENSGL- 408

Qy 299 ITKMLKATDLPKQDKAPDHGGLISKMLPGQDRGLCGELDQNLSCRCFKPHEKQKCOQAH 358

Db 409 --KRQVDFRVVLEETHDKRLNECIPFLE-----KYISKLNCSIV 451

Qy 359 SEDCDPVPALHTELDIAIRLVNVSNOQYGOILOMTRK--HLE 398

Db 452 VSQQTAPMKTKLLKQLI-----EQGENYKMTNLILHLD 486

Search completed: July 2, 2002, 11:44:34

Job time: 258 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:44:01 ; Search time 17.51 Seconds  
(without alignments)  
1094.585 Million cell updates/sec

Title: US-09-722-544A-2

Perfect score: 2632

Sequence: 1 MKIRAKENGRSRWQLHW.....FIGYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Match	Length	ID	Description
1	434	16.5	449	1	CLUS_HUMAN
2	416	15.8	451	1	CLUS_COTJA
3	413.5	15.7	447	1	CLUS_RABIT
4	410.5	15.6	445	1	CLUS_CANFA
5	403	15.3	448	1	CLUS_MOUSE
6	400.5	15.2	439	1	CLUS_BOVIN
7	400.5	15.2	446	1	CLUS_PIG
8	380	14.4	447	1	CLUS_RAT
9	168	6.4	191	1	CLUS_MESAU
10	151	5.7	1526	1	MY52_SCHPO
11	145.5	5.5	594	1	C1K1_YEAST
12	144.5	5.5	1938	1	MYHD_HUMAN
13	142.5	5.4	2683	1	CENE_HUMAN
14	141.5	5.4	3433	1	UTRO_HUMAN
15	141	5.4	1005	1	MY50_METJA
16	137	5.2	1940	1	MYH3_HUMAN
17	136	5.2	2230	1	GOG4_HUMAN
18	135	5.1	1060	1	EG51_XENLA
19	134.5	5.1	1957	1	YD86_SCHPO
20	134	5.1	5430	1	ACFY_HUMAN
21	132.5	5.0	852	1	RA50_THEMA
22	131.5	5.0	1679	1	Y109_YEAST
23	130.5	5.0	1940	1	MYH3_RAT
24	129.5	4.9	1427	1	REST_HUMAN
25	129.5	4.9	2077	1	TEGU_HSV6U
26	129	4.9	539	1	MY53_HYDAT
27	129	4.9	1853	1	MY5A_MOUSE
28	129	4.9	2704	1	BPAL_HUMAN
29	128.5	4.9	971	1	Y228_BORBU
30	128.5	4.9	1966	1	MY5B_CAEEL
31	128	4.9	1189	1	SCII_CHICK
32	128	4.9	1939	1	MYH1_HUMAN
33	128	4.9	2869	1	RBPI_PLAVB

34 127.5 4.8 1978 1 MYHB\_CHICK P10587 gallus gall  
35 127 4.8 793 1 KATA\_ARATH Q07970 arabidopsis  
36 127 4.8 1829 1 MY5A\_CHICK Q02440 gallus gall  
37 126 4.8 1941 1 MYH2\_HUMAN Q9UKX2 homo sapien  
38 126 4.8 1960 1 MYH9\_HUMAN P03579 homo sapien  
39 126 4.8 2418 1 SPCA\_HUMAN P38871 saccharomyc  
40 125.5 4.8 571 1 YH4\_YEAST P12844 caenorhabdi  
41 125.5 4.8 1969 1 MY5A\_CAEEL P49454 homo sapien  
42 125.5 4.8 3210 1 CENF\_HUMAN O67124 aquifex aeo  
43 125 4.7 978 1 RA50\_AQUAE P25386 saccharomyc  
44 125 4.7 1790 1 US01\_YEAST P29616 gallus gall  
45 124.5 4.7 1102 1 MYSC\_CHICK

#### ALIGNMENTS

RESULT 1  
CLUS\_HUMAN  
ID CLUS\_HUMAN STANDARD; PRT; 449 AA.  
AC P10909; P11380; P11381;  
DT 01-JUL-1989 (Rel. 11, Created)  
DT 01-JUL-1989 (Rel. 11, Last sequence update)  
DT 01-MAR-2002 (Rel. 41, Last annotation update)  
DE Clusterin precursor (Complement-associated protein SP-40,40)  
DE (Complement cytotoxicity inhibitor) (CLI) (NAL and NA2) (Apolipoprotein J) (Apo-J) (TRPM-2).  
GN CLU.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.  
RC TISSUE=Liver;  
RA Kirsbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,  
RA Hudson P., Walker I.D.;  
RT "Molecular cloning and characterization of the novel, human  
complement-associated protein, SP-40,40: a link between the  
complement and reproductive systems.";  
EMBO J. 8:711-718(1989).  
RN [2]  
RP SEQUENCE OF 2-449 FROM N.A.  
RX MEDLINE=89386692; PubMed=2780565;  
RA Jenne D.E., Tschopp J.;  
RT "Molecular structure and functional characterization of a human  
complement cytotoxicity inhibitor found in blood and seminal plasma:  
identity to sulfated glycoprotein 2, a constituent of rat testis  
fluid.";  
Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=90344779; PubMed=1974459;  
RA de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;  
RT "Apolipoprotein J: structure and tissue distribution.";  
Biochemistry 29:5380-5389(1990).  
RN [4]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=94237156; PubMed=8181474;  
RA Wong P., Taillefer D., Lalkins J., Pineault J., Chader G.,  
RA Tenniswood M.;  
RT "Molecular characterization of human TRPM-2/clusterin, a gene  
associated with sperm maturation, apoptosis and neurodegeneration.";  
Eur. J. Biochem. 221:917-925(1994).  
RN [5]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93186813; PubMed=7680346;  
RA Pineault J.M., Tenniswood M.;  
RT "Genomic organization and expression of the rat TRPM-2 (clusterin)  
gene, a gene implicated in apoptosis.";  
J. Biol. Chem. 268:5021-5031(1993).  
RN [6]

```

RP SEQUENCE FROM N.A.
RA Blechschmidt K., Rosenthal A.;
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.
RN [7]
RP SEQUENCE OF 61-449 FROM N.A.
RX MEDLINE-92020896; PubMed-1924317;
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,
RL Quirion R., Suh M.;
RN "Human gliomas and epileptic foci express high levels of a mRNA
RT related to rat testicular sulfated glycoprotein 2, a purported marker
RL of cell death.";
RN Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991);
RN [8]
RP SEQUENCE OF 140-449 FROM N.A.
RX MEDLINE-91365608; PubMed-3154963;
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,
RL Hochstrasser D.F., Pometta D., Merrill C.R.;
RN "HDL particle associated proteins in plasma and cerebrospinal fluid:
RT identification and partial sequencing.";
RL Appl. Theor. Electrophor. 1:73-76(1988).
RN [10]
RP SEQUENCE OF 23-33 AND 228-240.
RX MEDLINE-912330083; PubMed-1903084;
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,
RL Hochstrasser D.F.;
RN "Characterization of a human high density lipoprotein-associated
RT protein, NAL/NA2. Identity with SP-40,40, an inhibitor of complement-
RT mediated cytolysis.";
RL Arterioscler. Thromb. 11:645-652(1991).
RN [11]
RP SEQUENCE OF 23-52 AND 228-257.
RX MEDLINE-90354412; PubMed-2387851;
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,
RL Wetterau J.R., Busch S.J., Harmony J.A.K.;
RN "Purification and characterization of apolipoprotein J.";
RL J. Biol. Chem. 265:14292-14297(1990).
RN [12]
RP SEQUENCE OF 23-37 AND 228-242.
RX MEDLINE-90097955; PubMed-2601725;
RA Choi N.H., Mazda T., Tomita M.;
RN "A serum protein SP40,40 modulates the formation of membrane attack
RT complex of complement on erythrocytes.";
RL Mol. Immunol. 26:835-840(1989).
RN [13]
RP SEQUENCE OF 23-41 AND 228-246.
RX MEDLINE-93319521; PubMed-8328966;
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,
RN Wisniewski T., Frangione B.;
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is
RT complexed to SP-40,40 (apolipoprotein J), an inhibitor of the
RT complement membrane-attack complex.";
RL Biochem. J. 293:27-30(1993).
RN [14]
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-92201397; PubMed-1551440;
RA Kirszbaum L., Bozas S.E., Walker I.D.;
RN "SP-40,40, a protein involved in the control of the complement
RT pathway, possesses a unique array of disulphide bridges.";
RL FEBS Lett. 297:70-76(1992).
RN [15]
RP DISULFIDE BONDS.
RX MEDLINE-93147076; PubMed-1491011;
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;
RN "Identification of the disulfide bonds in human plasma protein
RT SP-40,40 (apolipoprotein-J)";
RL J. Biochem. 112:557-561(1992).
RN [16]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE-97477902; PubMed-9336835;
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,
RL Carr S.A., Crabb J.W.;
RN "Identification and characterization of glycosylation sites in human
RT serum clusterin.";
RN Protein Sci. 6:2120-2133(1997).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (APOPTOSIS)
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X14723; CAA32847.1; -
DR EMBL; M25915; AAA35692.1; ALT_INIT.
DR EMBL; J02908; AAA51765.1; ALT_INIT.
DR EMBL; M74816; AAA60321.1; -
DR EMBL; M64722; AAB06508.1; -
DR EMBL; L00974; AAA60567.1; -
DR EMBL; AF311103; -; NOT_ANNOTATED_CDS.
DR EMBL; M63379; AAB06507.1; -
DR EMBL; M63376; AAB06507.1; JOINED.
DR EMBL; M63377; AAB06507.1; JOINED.
DR EMBL; M63378; AAB06507.1; JOINED.
DR EMBL; A21577; CAA01560.1; -
DR PIR; A35833; A35833.
DR PIR; A37816; A37816.
DR PIR; B37816; B37816.
DR PIR; A41386; A41386.
DR PIR; PL0135; PL0135.
DR PIR; PL0136; PL0136.
DR PIR; S04662; S04662.
DR PIR; S07433; S07433.
DR PIR; S09339; S09339.
DR PIR; S34056; S34056.
DR PIR; S43646; S43646.
DR SWISS-2DPAGE; P10909; HUMAN.
DR MIM; 185430; -
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; Clb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;
FT SIGNAL.
FT CHAIN 1 22
FT CHAIN 23 449
FT CHAIN 23 227
FT CHAIN 228 449
FT DISULFID 102 313
FT DISULFID 113 305
FT DISULFID 116 302
FT DISULFID 121 295
FT DISULFID 129 285
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 291 291
FT CARBOHYD 354 354
FT CARBOHYD 374 374
FT CONFLICT 28 28
FT CONFLICT 47 47
FT CONFLICT 52 52
FT CONFLICT 305 305
CLUSTERIN.
BETA-CHAIN.
ALPHA-CHAIN.
INTERCHAIN.
INTERCHAIN.
INTERCHAIN.
INTERCHAIN.
INTERCHAIN.
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
N-LINKED (GLCNAC. . .).
D -> S (IN REF. 9 AND 10).
Q -> H (IN REF. 11).
G -> Q (IN REF. 11).
C -> M (IN REF. 10).

```

```

Query Match      16.5%  Score 434;  DB 1;  Length 449;
Best Local Similarity 25.7%;  Pred. No. 1.6e-18;
Matches 124;  Conservative 96;  Mismatches 208;  Indels 52;  Gaps 14;

Oy 30 MKPPLLVIVCLLWLKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 83
Db 1 MKTKLLIVGVLL-----TWESQVLDGTVSDNELQEMNOGSKYVKNKEIONAVN 51
Oy 84 GIKQMKIMMERKEHTNLMTLKKCKEOKEALKLLNEVOEHLEERLCRESLADSWG 143
Db 52 GVKQIKTLIETNEKRTLLSNLEBEAKKKEDALNETRESETKLKEPGVCNETMMALWE 111
Oy 144 ECRSCLENNCRITY-TCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLEE 202
Db 112 ECKPCLKQTKMFKYARVCRSGSLVGRQLEEFNLQSSPFYFWMNGDR-----IDSLLEN 165
Oy 203 DAOLTOEDV----FSQLTVDVNSLFNRSFNVFMQOEFTQSFHIS--DTOLTPEYP 257
Db 166 DRQOQTHMLVDQDHFSSRASSIIDELQDRF-----FTREPODQTY--HYLPFSLPHRRPHF 218
Oy 258 FPAFSKEPMTKADLE-OCWDIPNFFQLFNCFNSVSIYSEVSETITKMLKAIEDLPKQDKAP 316
Db 219 F--FPKSRIVSLMFPSPYEPFLNFMFPLEMIHEA-----QQAMDIHFHSPAP 267
Oy 317 DHGLISKMLPGQDRGLCGELQDNLSCFKPHEKCKQCAHLSDEC-----PDVPALHTEL 372
Db 268 QHPPTFEIREGDDRTVCREIRHNSGTCLRMKQDCCKREILSVDCSTNNPQAKLRREL 327
Oy 373 DEARLVNSVNOQYGOILQMTKRLKLEDTAYLVKMRGFGVHWSLANQAPETEIFNSIQ 432
Db 328 DESQVABRLTKNELLKQWKLMTSSLLQNLQFNQVNSLANLTOGEDQYILRVT 387
Oy 433 VVPRHIEGNISKQDETMMTDLISLPSSNFTLKIPLIESAESSNFIGYVAKALQHF-KEH 491
Db 388 TVAS-HTSDSDVPVSVTEVVVKLFDSDPITVTPVEVSRKNPKMEVVAEKALQYRKKH 446

RESULT 2
CLUS_CONJA
ID CLUS_CONJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN T64.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89239492; PubMed=2541393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calochy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
RT induced by different retroviral oncogenes in quail neuroretinal
RT cells.";
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
RT alternative promoters with distinct regulatory elements.";
RL Eur. J. Biochem. 229:215-223(1995).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER

```

```

(BY SIMILARITY).
-1- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.
-1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
-----
THIS SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see http://www.isb-sib.ch/announce/
or send an email to license@sib-sib.ch).
-----
EMBL; X15825; CAA33823.1; -
EMBL; X80760; CAA56733.1; -
PIR; S07714; S07714.
InterPro; IPR000753; Clusterin.
Pfam; PF01093; Clusterin; 1.
SMART; SM00035; Cla; 1.
SMART; SM00030; CLB; 1.
PROSITE; PS00492; CLUSTERIN_1; 1.
PROSITE; PS00493; CLUSTERIN_2; 1.
Glycoprotein; Signal.
SIGNAL 1 18 POTENTIAL.
CHAIN 19 451 CLUSTERIN.
FT CHAIN 19 230 BETA-CHAIN (SMALL SUBUNIT) (POTENTIAL).
FT CHAIN 231 451 ALPHA-CHAIN (LARGE SUBUNIT) (POTENTIAL).
FT DISULFID 98 314 INTERCHAIN (BY SIMILARITY).
FT DISULFID 109 306 INTERCHAIN (BY SIMILARITY).
FT DISULFID 112 303 INTERCHAIN (BY SIMILARITY).
FT DISULFID 117 296 INTERCHAIN (BY SIMILARITY).
FT DISULFID 125 286 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 99 99 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 141 141 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 278 278 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 355 355 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CARBOHYD 375 375 N-LINKED (GLCNAC. ) (POTENTIAL).
FT CONFLICT 437 437 A -> P (IN REF. 1).
FT CONFLICT 445 451 KONNTIE -> SRTTP (IN REF. 1).
SQ SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;

Query Match      15.8%  Score 416;  DB 1;  Length 451;
Best Local Similarity 25.7%;  Pred. No. 1.9e-18;
Matches 118;  Conservative 93;  Mismatches 180;  Indels 68;  Gaps 15;

Oy 63 LKSFSEVGEIDAEVKKALTGIKIMMERKEHTNLMSTLKKCKEOKEALKLLNE 122
Db 27 LKQLSAGSKYIDAEVENAINGVKMTLMDKTSKEHQAMLHLEETKKKKEAVKLALE 86
Oy 123 VQEHLEEEERLCRESLADSWGECRSCLENNCRITY-TCQPSWSSVKNKIERFRKIYQF 181
Db 87 KEKQAEKQEVNCTMLSLWECKPCLKHTCMRVYSKMHSGSLVGRQLEEFNRSPPF 146
Oy 182 LFPFHEDNEKDLPISEKLEE-----DAQLTOMEDVFSQLTVDVNSLFNRSFNVFMQOQ 236
Db 147 SIWNGRIRIDDLDRQQRERRFEDLEERFGLMEDVEDIFQDSTQLYGAFPPFR---- 202
Oy 237 EFDQTFQSHEFISDTDLTPY--FPPAFSKBPMTKADLEKQCDIPN-----FFOLF 285
Db 203 -----TPPFGGFREAF-VPPQVRVHL-----VPRRLSRELHFFHFFOHPM 240
Oy 286 NFSVSIYSEVSETITKMLKAIEDLPKQDKAPDH--GGLISKMLP-GODRGLCGELQDNL 342
Db 241 HGPHRLFQPLFEMTOHMLDG-----GHGAWEHPLGFGFATESRNFSTDMVRCREIRNSA 294
Oy 343 RCFKFEKCKQCAHLSDEC-----PDVPALHTELDEAIRLVNSVNOQYGOILQMTKRLK 398
Db 295 GCLMRDECEKCEKREILAVDCSQTDPVQSQREQFEDALRAERFTRFDLLLSAFAEML 354
Oy 399 DTAYLVKMRGFGVHWSLANQAP-----ETEIFNSIQVVPRIHSGNTSKQDETMMTD 452
Db 355 NTSSLLDQLNRQFGVHWSRLNLTQGNDFLQVTVFVSK-----TPMLEDP--SAPADQVTV 408

```

QY 453 LSLPSSNTLKIPLSEAESNFIYVYVAKALQHFKEH 491  
DB 409 VQLFDSEPLSLVPGDISWDDPRFMEIVAEQALQHYKQN 447

RESULT 3  
CLUS\_RABIT STANDARD; PRT; 447 AA.  
AC Q9XSC5;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Clusterin precursor (Apolipoprotein J) (Apo-J).  
GN CLU.  
OS Oryctolagus cuniculus (Rabbit).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
OX NCBI\_TaxID=9986;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-JAPANESE WHITE;  
RA Miyata M., Biro S., Kaieda H., Eto H., Orihara K., Kihara T.,  
Obata H., Matsushita N., Matsuyama T., Tei C.;  
RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells  
by balloon injury.";  
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH (BY SIMILARITY).  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; AF118852; RAD24461.1; -;  
DR InterPro; IPR000753; Clusterin; 1.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; CLa; 1.  
DR SMART; SM00030; Clb; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
KW Glycoprotein; Plasma; Signal.  
FT SIGNAL 1 22 BY SIMILARITY.  
FT CHAIN 23 447 CLUSTERIN.  
FT CHAIN 23 225 BETA-CHAIN (A CHAIN) (BY SIMILARITY).  
FT CHAIN 226 447 ALPHA-CHAIN (B CHAIN) (BY SIMILARITY).  
FT DISULFID 101 311 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 112 303 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 115 300 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 120 293 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 128 283 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 85 85 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 289 289 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 326 326 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 352 352 N-LINKED (GLCNAC. .) (POTENTIAL).  
FT CARBOHYD 372 372 N-LINKED (GLCNAC. .) (POTENTIAL).  
SQ SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A CRC64;

Query Match 15.7%; Score 413.5; DB 1; Length 447;  
Best Local Similarity 24.1%; Pred. No. 2.7e-18;  
Matches 111; Conservative 93; Mismatches 204;  
QY 52 TWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKQKIMMERKEHTNLMST 105

DB 13 SWERGVLGDQLVSNELQEMSTQSGKYIDREIQNAVKGVQIKTLIEKTNERTLLSV 72  
QY 106 LKCKREEKOEALKLLNEVOEHLLEERLCRESLADSWGECSCLENNCMRIYT-TCOPSM 164  
DB 73 LEEAKNKEDALNETRDSKTKAPFVCNMTMMALWECKPCLKOTCMKFYARVCRSGS 132  
QY 165 SSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIETEDAQLTQMBDFVSQLTVDVNSLF 224  
DB 133 GLVGRQLEELNQSPFFYFWINGDR-----IDSLLENDRQOQSHVLVD-----MQDSF 179  
QY 225 NRSFNVRQMQQEFQTFQSHFISDTLDTEPYFFFAFSPKPEMTKADLEOCWDIP----- 278  
DB 180 NRATGI---MDELFDQDFRFTHRPQDTFYHSPESY---FRPPFLHYAKSRILVRNIMPLSLYG 234  
QY 279 --NFTQFCNFSVSVSVSEITITMKLAIEDLPKQD---KAPDHGGLISKMLPGQDRLG 333  
DB 235 PLNFQDMQFPFFEMIHQAQAMDVHLHSPAYTPNVEFTGGPD-----DRAV 282  
QY 334 CGELDNLSRCFKFHEKCKQCAHLSEDC-----PDVPALHTELDEAIRLVNYSNQYQGI 389  
DB 283 CKEIHRNSTGCLRMKDKQCAKQCEILLVDCSANNPSONQLRQELNDSLRLAEELTKRYNEL 342  
QY 390 LQMKRKHLEDYALVEKMRGQFGWSELANQAPETETIENSQVVPRIEHNISQDETM 449  
DB 343 LOSYQWKMLNTSLLDQNPNEQFNWVSQLANLTQGPQYLYRVSTVTS-HTSESEAPSRVT 401  
QY 450 MTDLSLPSNSTLKIPLSEAESNFIYVYVAKALQHFKE 490  
DB 402 EVVVKLFDSDPITITPEVSRDNPFKMETVAEKALQYERK 442  
RESULT 4  
CLUS\_CANFA STANDARD; PRT; 445 AA.  
AC P25473;  
DT 01-MAY-1992 (Rel. 22, Created)  
DT 01-MAY-1992 (Rel. 22, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Clusterin precursor (Glycoprotein 80) (GP80).  
GN CLU.  
OS Canis familiaris (Dog).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
OX NCBI\_TaxID=9615;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-91236776; PubMed-2033078;  
RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pillarsky C.,  
RA Appel D., Haase W., Mann K., Weller A., Koch-Brandt C.;  
RT "Molecular cloning of gp 80, a glycoprotein complex secreted by  
kidney cells in vitro and in vivo. A link to the reproductive system  
and to the complement cascade.";  
RL J. Biol. Chem. 266:9924-9931(1991).  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH.  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M55251; AAA30846.1; -;  
DR PIR; A40018; A40018.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.





```

FT CHAIN 227 447 (BY SIMILARITY)
FT ALPHA-CHAIN (LARGE SUBUNIT)
FT (BY SIMILARITY)
FT INTERCHAIN (BY SIMILARITY)
FT DISULFID 101 312
FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY)
FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY)
FT DISULFID 120 284 INTERCHAIN (BY SIMILARITY)
FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY)
FT CARBOHYD 102 102 N-LINKED (GLCNAC. .) (PROBABLE)
FT CARBOHYD 144 144 N-LINKED (GLCNAC. .) (PROBABLE)
FT CARBOHYD 290 290 N-LINKED (GLCNAC. .) (PROBABLE)
FT CARBOHYD 327 327 N-LINKED (GLCNAC. .) (PROBABLE)
FT CARBOHYD 353 353 N-LINKED (GLCNAC. .) (PROBABLE)
FT CARBOHYD 373 373 N-LINKED (GLCNAC. .) (PROBABLE)
FT CONFLICT 10 10 L -> M (IN REF. 4)
FT CONFLICT 335 336 K -> TV (IN REF. 3)
FT CONFLICT 350 350 K -> N (IN REF. 3)
SQ SEQUENCE 448 AA; 51655 MW; A860600A6F8D47F6 CRC64;

Query Match 15.3%; Score 403; DB 1; Length 448;
Best Local Similarity 23.4%; Pred. No. 1.2e-17;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

QY 34 LTVFIWLLWLDKSHCAPTKDKTALSENKLSFSEVEIDA-----DEEVKALTGIKO 87
DB 3 ILLLCVALLI-----WDNGMVLGEQVSDNEIQELSTQGSRYINKEIQNAVQGVKH 54
QY 88 MKTMRKEKEHTNLMSTLTKCKREEQKALLLNEVQHELEERLCRESLADSGECRS 147
DB 55 IKTLIEKTAERKSLNSLEEAKKKEDALEDFRSEMKLKAPFVCNETMMALWEECKP 114
QY 148 CLNNCHRIYT-TCQPSWSVKKNIERFRKYYQFLFPFHEDNEKDLPISEKLIEDAOL 206
DB 115 CLKHTCMKFYARVCRSGSLVGQOLEEFLNQSPFFYFWMGDR-----IDSLLESDRQ 168
QY 207 TQ-----MEDVFSOLTVDVNSLENRSNVFRMQQEDFTQSFHFIISDTDLTEPYEFP-- 259
DB 169 SQVLDAMQDSFARAGIITFLFDOR--FAR-----ELHDPIYS-----PIGFPHKR 214
QY 260 ---AFSKEPTKADLEQ-WDIPNFFQLFCNFSVSVESSETITMKLKAIEDLPKQOKA 315
DB 215 PHELYPKSLRVLSMSPSHYGPFSFHMFPFFEMIHQAQAMQVQLHSPAFQFPD-- 272
QY 316 PDHGLLISMLPGQ-DRGLGELDQNLSCFKFHEKQCKQAHLSDC-----PDVFAHPT 370
DB 273 -----FLREGEDDRTVCKEIRRNSTGCLKMKGQCKEQEILSYDCSTNNPAQANLRQ 324
QY 371 ELDEAIRLVNVSNOQGOILQMTKHLKEDTAYLVEKMRGQFGVSELANQAPETEIFNS 430
DB 325 ELNDSQVABRLTEQYKELLQSQSKMLNTSSLEQLNDQFNWVSQLANTQGEDKYILR 384
QY 431 IQVVPRIHEGNIKQDETMTDLISLPSSNFTLKIPLEESAESSNFTGYVAKALQHKF 489
DB 385 VSTV-ITHSSDSEVPSTRVTEVVVKLFDSDPITVVLPEVSKDNPKFMDTVAEKALQEYR 442

RESULT 6
CLUS_BOVIN STANDARD; PRT; 439 AA.
AC PI7697;
DT 01-AUG-1990 (Rel. 15, Created)
DT 01-AUG-1990 (Rel. 15, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE Clusterin precursor (Glycoprotein III) (GP111).
GN CLU.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

```

```

RC TISSUE=Chromaffin granules;
RA MEDLINE=9021681; PubMed=1691174;
RX Palmer D.J., Christie D.L.;
RT "The primary structure of glycoprotein III from bovine adrenal
RT medullary chromaffin granules. Sequence similarity with human serum
RT protein-40.40 and rat Sertoli cell glycoprotein.";
RL J. Biol. Chem. 265:6617-6623(1990).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: J05391; AAA30554.1; -
DR PIR: A35744; A35744. Clusterin.
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; CLU; 1.
DR SMART: SM00030; CLU; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 439 CLUSTERIN.
FT CHAIN 23 439 BETA-CHAIN (A CHAIN).
FT CHAIN 222 439 ALPHA-CHAIN (B CHAIN).
FT DISULFID 96 305 INTERCHAIN (BY SIMILARITY).
FT DISULFID 107 287 INTERCHAIN (BY SIMILARITY).
FT DISULFID 110 294 INTERCHAIN (BY SIMILARITY).
FT DISULFID 115 287 INTERCHAIN (BY SIMILARITY).
FT DISULFID 123 277 INTERCHAIN (BY SIMILARITY).
FT CARBOHYD 80 80 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 97 97 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 139 139 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 283 283 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 320 320 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 346 346 N-LINKED (GLCNAC. .) (POTENTIAL).
FT CARBOHYD 366 366 N-LINKED (GLCNAC. .) (POTENTIAL).
SQ SEQUENCE 439 AA; 51114 MW; 36979481A61B02DE CRC64;

Query Match 15.2%; Score 400.5; DB 1; Length 439;
Best Local Similarity 24.5%; Pred. No. 1.6e-17;
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 52 TWKDKTAISE-NLKSFSEVEIDADEVKALTGIQKIMMERKEKEHTNLMSTLKKCR 110
DB 13 SWESGWAISDKELQEMSTEGSKYVNKEIKALKEVQIKTQIBOTNEERKLLLSLEAK 72
QY 111 EEKQEAALLNEVQHELEERLCRESLADSGECRSCLNNCMRIYT-TCQPSWSVKN 169
DB 73 KKKEDALNDRDSENKLSAQGVNCTMTALWEECKPCLKQTCMKFYARVCRSGSLGVGH 132
QY 170 KIERFFRKYYQFLFPFHEDNEKDLPISEKLIEDAQITQMEDVFSQITVDVNSLNFNSFN 229
DB 133 QLEEFNLQNSPPFYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRASS 179
QY 230 VFQMQQEFQDTQSHFI-----SDTDLTEPY-----FF----PAFSKEPTKADLEQC 274
DB 180 I-----MDELQDFRFFLLRRPDQTYQYSPFSFGRSUFFNPKSFARVMVFPFLEPF 232
QY 275 WDIPNFFQLFCNFSVSVISYESTITMKLKAIEDLPKQDKAPDHGGLISMLPGQDRGLC 334

```

FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	290	290	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. . .) (POTENTIAL) .
QY	SEQUENCE	446 AA;	51774 MW;	B1D5B434B668E3AA CRC64;
Query Match 15.2%; Score 400.5; DB 1; Length 446;				
Best Local Similarity 24.4%; Pred. No. 1.7e-17;				
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps				
QY	34	LLVFIYVCLLMKLSHCAPTWK-----DKTASENLKFSEVGEIDADEVVKALTGK	86	
DB	4	LLLLVLGILL-----TWENGPFVLGDKRAISDKELQEMTEGSGYVKNKEIKNALKEYK	54	
QY	87	QMKIMMERKEHTNLMSTLLKKCEEKQEAALKLLNEVQEHLEERLCRESLADSGECR	146	
DB	55	QIKTLIEQSNEERKSLSSLEEAKKKEDALNTRTETKLKGSQGLCNETMMALWEECK	114	
QY	147	SCLENNCMRYIT-TCOPSWSGSVKNKIERFRPKIYQFLFPHEHONKDLPTSEKLIIEDAQ	205	
DB	115	PLQKOTCMKFYARVCRSGSLGVGHQLEEFNLQSSPFYFWINGDR-----IDSLMENDRQ	168	
QY	206	LTQMEDVFSQLTVDVNSLFRNSRVFMQOMQEEFDQTFQSHFIS-----DIDLTEPY	256	
DB	169	QSHVMDI-----MEDSFNRASNI-----MDLFDQRRFNREPDFQTFSPFGSSHR	214	
QY	257	---FFPAFSKEPWTKADLEOCWDIPNPFQLFONFSVSIYESVETITKMLKAIEDLPKOD	313	
DB	215	GSLEFPNPKSRFARNIPFPLFTDL-NYHDMFQPF---FDMIHQAQAMDAHLHRIYPH-	268	
QY	314	KAPDHGGLISKMLP--GQDRGLGCLGELDONLSRCFKFHEKCKQCOAHLSEDC----	367	
DB	269	-FPAG-----VPENENDRAVCKEIRHNTGCLRKDKOCEKREILSVDCSASNSQM	321	
QY	368	LHTELDEAIRLVNYSNOQYQGIQMTKKHLEDYAILVEKMRGFGWVSELANOQAPETEI	427	
DB	322	LRQELYSLOMAEKFSKLYDOLLQSYOQKMLNLSLLKQLNEQFSWVSYQLANLTQNDRY	381	
QY	428	FNSIQVPRITHEGNISKQDETMTDLSILPSSNFTLKIPLEESAESSNFTGIYVAKALOH	487	
DB	382	YLQVTTV-NSHGSDPSVPISGLTKVVKVLFDSYPITLIPOEVS--DPKFMETVAEEALQO	438	
QY	488	FKE 490		
DB	439	YRQ 441		
RESULT 8				
CLUST_RAT	CLUST_RAT	STANDARD;	PRT;	447 AA.
AC	P05371;			
DT	01-NOV-1988 (Rel. 09, Created)			
DT	01-FEB-1994 (Rel. 28, Last sequence update)			
DT	01-OCT-1996 (Rel. 34, Last annotation update)			
DE	Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM 2).			
DE	2).			
GN	CLU.			
OS	Rattus norvegicus (Rat).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.			
OX	NCBI_TaxID=10116;			
RP	[1]			
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.			
RX	MEDLINE=88000523; PubMed=3651384;			
RA	Collard M.W., Griswold M.D.;			
FT	"Biosynthesis and molecular cloning of sulfated glycoprotein 2			
FT	secreted by rat Sertoli cells."			
RL	Biochemistry 26:3297-3303(1987).			
RP	[2]			
RP	SEQUENCE FROM N.A.			

RA Pineault J.M., Tenniswood M.;  
RN Submitted (XXX-1991) to the EMBL/GenBank/DBJ databases.  
RL [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89149740; PubMed-2920020;  
RA Bettuzzi S., Hilpakka R.A., Gilna P., Liao S.;  
RT "Identification of an androgen-repressed mRNA in rat ventral prostate  
as coding for sulphated glycoprotein 2 by cDNA cloning and sequence  
analysis.";  
RL Biochem. J. 257:293-296(1989).  
RN [4]  
RP SEQUENCE OF 22-51 AND 227-256.  
RX MEDLINE-88326333; PubMed-3415696;  
RA Cheng C.-Y., Chen C.C., Feng Z., Marshall A., Bardin C.W.;  
RT "Rat clusterin isolated from primary Sertoli cell-enriched culture  
medium is sulfated glycoprotein-2 (SGP-2).";  
RL Biochem. Biophys. Res. Commun. 155:398-404(1988).  
RN [5]  
RP CHARACTERIZATION OF TRPM-2.  
RX MEDLINE-90134121; PubMed-2299741;  
RA Bandyk M.G., Sawczuk I.S., Olsson C.A., Katz A.E., Buttyan R.;  
RT "Characterization of the products of a gene expressed during  
androgen-programmed cell death and their potential use as a marker  
of urogenital injury.";  
RL J. Urol. 143:407-413(1990).  
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH.  
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -1- DEVELOPMENTAL STAGE: EXPRESSED BY CELLS UNDERGOING PROGRAMMED  
CC DEATH AS A RESULT OF THE HORMONAL STIMULI OR A TRAUMATIC INSULT.  
CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED  
CC CARBOHYDRATES.  
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC EMBL; M16975; AAA41273.1; -  
CC EMBL; M64723; AAA42298.1; -  
CC EMBL; M64733; AAA42299.1; -  
CC EMBL; X13231; CAA31618.1; -  
CC PIR; A27205; A27205.  
CC PIR; A31575; A31575.  
CC PIR; B31575; B31575.  
CC PIR; A45890; A45890.  
CC PIR; S18491; S18491.  
CC InterPro: IPRO00753; Clusterin.  
CC Pfam: PF01093; Clusterin; 1.  
CC SMART: SM00035; Cla; 1.  
CC SMART: SM00030; Clb; 1.  
CC PROSITE: PS00492; CLUSTERIN\_1; 1.  
CC PROSITE: PS00493; CLUSTERIN\_2; 1.  
KW Sulfation; Glycoprotein; Spermatogenesis; Signal.  
FT SIGNAL 1 21  
FT CHAIN 22 447 CLUSTERIN.  
FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT).  
FT CHAIN 227 447 ALPHA-CHAIN (LARGE SUBUNIT).  
FT DISULFID 101 312 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 120 294 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CONFLICT 187 187 D -> H (IN REF. 1).  
SQ SEQUENCE 447 AA; 51375 MW; 9E2FA33E5E0C146E CRC64;  
  
Query Match 14.4%; Score 380; DB 1; Length 447;  
Best Local Similarity 23.5%; Pred. No. 2.9e-16;  
Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;  
  
QY 34 LLVFIVCLLWLDKSHCAPTWKDKTAISEMLKSFSEVEID-----ADEVVKALTGIIQ 87  
DB 4 LLLCVALL-----TWDMGVLGEQEFSDNEIQELSTOGSRYVNEIIONAVOGV 54  
QY 88 MKIMMERKEKEHTNLMSTLKKCKREEKQEAALKLLNEVQEHLEEEERLCRESLADSWCECRS 147  
DB 55 IKTLIEKTNAERKSLNLSLEAKKKEGALDDTRDSEMMLKAPPEVCNCTMMLWEECKP 114  
QY 148 CLENNCMRIYT-TCQPSWSSVKNKIERFRKIYQVLFPPHEDNEKDLPISEKLIIEDAQ 206  
DB 115 CLKHTCMKFYARVCRSGSLVGRQLEEFNLQSSPFYFMNGDR-----IDSLLESDRQQ 168  
QY 207 TQ----MEDVFQSLTVDVNSLENRSFNVRQMOQEFDTQFSHFISDTDLTEPYFPF--- 259  
DB 169 SQVLDAQDSFTRASGIIDTLFQDRF--FTHPEQDI-----HHF-----SPMGPPHXR 214  
QY 260 ---AFSKEPMTKADLQECWDIP-NFFQLFCNFSVSIVSEVSETITMKLKAIEDLPKQDKA 315  
DB 215 PHFLYPKSLRSLMPLSHYGPLSPHNMFPFFDMLHQAQAMDVLHSPALQFPDVOFL 274  
QY 316 PDHGLISLMLPQO-DRGLCGELDONLSRCFHEKCKQCAHLSDC-----PDVPAHIT 370  
DB 275 KE-----GEDDPTVCYKEIRHNSGTCLMKMGQCEKCEITLSVDCSTNNPAQANLRQ 324  
QY 371 ELDEAIRLVNSNOQYQIQLMTRKHLEDTAYLVKMRGQFGWVSELAN--QAPETEIIIF 428  
DB 325 ELNDSLOVAERLTQQYNELHLSLOKMLNTSSLLLEQNDQFTWVSQLANLTGDDQYLRV 384  
QY 429 NSIQVVPRIHEGNISKQDQETMTDLSILPSSNFTKIPLEESAESSNFYGVYAKALQHF 488  
DB 385 STVTT----HSSDSEVPSSRVTEVVVKLFDSPTTVVLPEEVSKDNPKFMDTVAERAKQEV 440  
QY 489 K 489  
DB 441 R 441

## RESULT 9

CLUS\_MESAU  
ID CLUS\_MESAU STANDARD; PRT; 191 AA.  
AC P14863;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Clusterin (Sulfated glycoprotein 2) (SGP-2) (Fragment).  
GN CLU.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OX NCBI\_TaxID=10036;  
RN [1]  
RP SEQUENCE OF 24-123 FROM N.A.  
RX MEDLINE-89386721; PubMed-2780570;  
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;  
RT "Changes in brain gene expression shared by scrapie and Alzheimer  
disease.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).  
RN [2]  
RP SEQUENCE OF 1-23 AND 124-191 FROM N.A.  
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;  
RL Submitted (AUG-1989) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY

```
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH.  
CC  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED  
CC CARBOHYDRATES.  
CC  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; M26640; AAA37102.1; -;  
CC InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; Cla; 1.  
DR PROSITE; PS00492; CLUSTERIN_1; PARTIAL.  
DR PROSITE; PS00493; CLUSTERIN_2; 1.  
KW Sulfation; Glycoprotein; Spermatogenesis.  
FT NON_TER 1  
FT CHAIN 1  
FT CHAIN <1 >191 CLUSTERIN.  
FT CHAIN <1 >15 BETA-CHAIN (SMALL SUBUNIT).  
FT CHAIN 16 >191 ALPHA-CHAIN (LARGE SUBUNIT).  
FT CARBOHYD 79 79 N-LINKED (GLCNAC. .) (PROBABLE).  
FT CARBOHYD 116 116 N-LINKED (GLCNAC. .) (PROBABLE).  
FT NON_TER 191 191  
SQ SEQUENCE 191 AA; 22099 MW; 20A9E7E66963C619 CRC64;  
  
Query Match 6.4%; Score 168; DB 1; Length 191;  
Best Local Similarity 35.1%; Pred. No. 0.00075;  
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps 1;  
  
QY 330 DRGLGELDONLSRCFKHEKCKQKQAHLSDEC-----PDVPALHTELDEAIRLVNSNQ 385  
DB 69 DRAVCKEIRHNSGTGLKMKGQCEKQCEILSVDCSANNPAQAHLRQELNDSLQVAERLTOR 128  
  
QY 386 YGOILQMTKRKHLEDYALVEKMRGQFGWVSELAN 419  
DB 129 YNELLSLQTKMLNTSSLEQLNEQFNWVSQLAN 162  
  
RESULT 10  
MYS2_SCHPO STANDARD; PRT; 1526 AA.  
AC Q9USI6; p78969;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE Myosin type II heavy chain 1.  
GN MYO2 OR SPC645.05C  
OS Schizosaccharomyces pombe (Fission yeast).  
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;  
CC Schizosaccharomycetales; Schizosaccharomycetaceae;  
CC Schizosaccharomycetes.  
OX NCBI_TaxID=4896;  
RN [1]  
RP SEQUENCE FROM N.A., AND FUNCTION.  
RC STRAIN=972;  
RX MEDLINE=98075862; PubMed=9415380;  
RA May K.M., Watts F.Z., Jones N., Hyams J.S.;  
RT "Type II myosin involved in cytokinesis in the fission yeast,  
RT Schizosaccharomyces pombe";  
RL Cell Motil. Cytoskeleton 38:385-396(1997).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=972;  
RA Wood V., Rajandream M.A., Barrell B.G., Rieger M.;  
RL Submitted (MAR-1999) to the EMBL/GenBank/DBJ databases.
```

```
CC  
CC -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE  
CC CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.  
CC MAY WORK IN CONJUNCTION WITH MYO3.  
CC  
CC -!- SUBUNIT: BINDS TO CDC4 AND RLC1.  
CC -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.  
CC  
CC -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
CC  
CC EMBL; U75357; AAC49908.1; -;  
CC EMBL; AL049498; CAB39901.1; -;  
CC HSSP; P08799; LMND.  
DR InterPro; IPR000048; IQ.  
DR InterPro; IPR004009; Myosin_N.  
DR InterPro; IPR001609; myosin_head.  
DR Pfam; PF00612; IQ; 1.  
DR Pfam; PF00063; myosin_head; 1.  
DR Pfam; PF02736; Myosin_N; 1.  
DR PRINTS; PR00193; MYOSINHEAVY.  
DR PRODOM; PD000355; myosin_head; 1.  
DR SMART; SM00015; IQ; 1.  
DR SMART; SM00242; MYSC; 1.  
DR PROSITE; PS00096; IQ; 1.  
KW Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;  
KW Alkylation.  
FT DOMAIN 1 757 MYOSIN HEAD-LIKE.  
FT DOMAIN 758 787 IQ.  
FT DOMAIN 787 1244 COILED COIL (POTENTIAL).  
FT NP_BIND 170 177 ATP (POTENTIAL).  
FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).  
FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).  
FT CONFLICT 1337 1337 S->R (IN REF. 1).  
SQ SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;  
  
Query Match 5.7%; Score 151; DB 1; Length 1526;  
Best Local Similarity 20.7%; Pred. No. 0.089;  
Matches 93; Conservative 81; Mismatches 221; Indels 54; Gaps 15;  
  
QY 65 SFSEVGELDADEEVKKGALTGKIMMERKEKEHTNLMSTLKKREKQBAKLLNEVQ 124  
DB 895 SFSETKQ-QNENLQRESASLKQINNELESELEKTSKVETL-----LSEQNELK 942  
QY 125 EHLEEEERLCRESLADSGGCRSCLENNC-----MRIYTCOPSWSVKNK---IERFF 175  
DB 943 EKLSELEK-----DLLDTKGELESRENNNAVLSKAEFNEQCKSLQETIVTKDELDKLT 998  
QY 176 RKIYQFLFPFHEDNEKDLPISEKLEIEADAQLTQMEDVFSQLTVDVNSLFNRSFNVFRQM 235  
DB 999 KYISDYKTEIQEMRLTNQKNKESIQEGSLSESILKRVKKLERE-NSTLSDVSLKQK 1057  
QY 236 QEPD-QTFQSHFISDTDLTEPYFPFAPFAPSKADLEQCDWDPNPFQFCNFSVIYE 293  
DB 1058 EELSVLKGVOELTINNLEEKVNYLEADVQKLPKLKKELESLNDKDLQYQLQATKNKELEA 1117  
QY 294 SVSETITKMLKAIEDLP-KODKAPDGHGLISKMLPGDGRGLCGELDONL----SRCKFFH 348  
DB 1118 KVRECLANNIKSLKELENKEEKQNLSADSKYIELQ-----EIHENLLKVSLENYK 1171  
QY 349 EKQCKQAHLSDECDDVPALHTELDEAIRLVNSNQ-----QYGQILQMTKRHLEDTA 401  
DB 1172 KKYBGLQLDL-EGLKQVDVTFNQELSKKKHRDLTFNHESLLRQSASYKEKLSASSENKDL 1230  
QY 402 YLVEKMRGQFGWVSELANQAPETEIFNSIQVPRPIHE-GNIKQDETMTDLSILPSSN 460  
DB 1231 NKVSSLTQKVNELSPKASKVPELE-----RKITNLMHEYSQLGTFDEKRRKALIASRDN 1285
```



```
DR InterPro: IPR001609; myosin_head.  
DR Pfam: PF00612; IQ; 2.  
DR Pfam: PF00063; myosin_head; 1.  
DR Pfam: PF02736; Myosin_N; 1.  
DR Pfam: PF01576; Myosin_tail; 1.  
DR PRINTS: PR00193; MYOSINHEAVY.  
DR PRODOM: PD000355; myosin_head; 1.  
DR SMART: SM00015; IQ; 1.  
DR SMART: SM00242; MYSC; 1.  
DR PROSITE: PS00096; IQ; 1.  
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;  
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;  
KW Multigene family.  
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.  
FT DOMAIN 785 814 IO.  
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).  
FT NP_BIND 179 186 ATP (POTENTIAL).  
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).  
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).  
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).  
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).  
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).  
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;  
  
Query Match 5.5%; Score 144.5; DB 1; Length 1938;  
Best Local Similarity 20.7%; Pred. No. 0.29;  
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;  
  
QY 59 ISENKASFSEVGEIDADEV-KKALTIKQKIMMERKEHEHNLSTLKKCR-----EEK 113  
DB 1150 IIRLEAS--GATSAQIEMKREAEFKMRDRLEATLQHEATAATLKKQADSVAEL 1207  
QY 114 QEALKLNEVQHELEERLCRESLAD--SWGECRSLNMCNRIYVTCOPSSVKNKI 171  
DB 1208 GEQIDNLRVQKLEKEKSELKMEIDMASNIEALSKSNIERTCTVDEQSEIKAKD 1267  
QY 172 ERFRKTYQFLPPHEDNEKDLPISEKLIBEDAQLTQMEDVFSQTVVDVNSLFRNSPNVF 231  
DB 1268 EQQTQLI-----HDLNMOKARLOTQNGELSHRVEEKESLSQTSKQALTOQLBEELK 1320  
QY 232 ROMQOEFD-QTFQSHFIS---DIDLTPYFFPAFSPKPTKADLEQCV-----DIPNFF 281  
DB 1321 ROMEETAKNAMAHLQSSRHDCDLLR---EQYEEESQAEALQALSKANSEVAQWK 1376  
QY 282 QLFNCFVSIVSEYSETITKMLKAIEDL-PRQDKAPDHGGLISK---MLPQDQRGLACGEL 337  
DB 1377 TKYETDAIQRTEELEAKKLAQRLOEAEEKTETANSKCSLEKTKQRLQGEVEDLMRDL 1436  
QY 338 DQNLsrc-----FKPHEKQKQCAHLSEDCPDVPAHTELDEAIRLVNVS 383  
DB 1437 ERSHTACATLDKKORNFQKVLAEWKQKLDSEQAEEAAQKESRSLSTEL---FKMRNAYE 1493  
QY 384 QYQGI--LQWTRKHLEDATVYVEKMRGQGVWSELANOAPET-----EII 427  
DB 1494 EVDQLETLRNKNLQEE-----ISDLTEQIAETGKNLOEAETKKLVQEE 1540  
QY 428 FNSIQVVPRIHEGNISKQDETM---TDLSPSSNFTLKIPLSESAESSNFIGYVYVAKA 484  
DB 1541 KSDLOVALEEVGSLHEESKILRVQLELSQVKSSELDKRVIEKDEETEQLKRNQSQAEEA 1600  
QY 485 LQ 486  
DB 1601 LQ 1602  
  
RESULT 13  
CENE_HUMAN STANDARD; PRT; 2663 AA.  
AC Q02224;  
DT 01-JUL-1993 (Rel. 26, Created)  
DT 01-JUL-1993 (Rel. 26, Last sequence update)  
DT 30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE Centromeric protein E (CENP-E protein).  
GN CENPE.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
OX NCBI_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RX MEDLINE=93024922; PubMed=1406971;  
RA Yen T.J., Li G., Schaar B.T., Sziliak L., Cleveland D.W.;  
RT "CENP-E is a putative kinetochore motor that accumulates just before  
RL mitosis."; Nature 359:536-539(1992).  
RL Nature 359:536-539(1992).  
RN [2]  
RP CHARACTERIZATION.  
RX MEDLINE=95196755; PubMed=7889940;  
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
RL microtubule motor."; EMBO J. 14:918-926(1995).  
RN [3]  
RP CHARACTERIZATION.  
RX MEDLINE=98437347; PubMed=9763420;  
RA Chan G.K.T., Schaar B.T., Yen T.J.;  
RT "Characterization of the kinetochore binding domain of CENP-E reveals  
RL interactions with the kinetochore proteins CENP-F and hBUBR1."; J. Cell Biol. 143:49-63(1998).  
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
CC KINETOCHORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
CC AND/OR SPINDLE ELONGATION.  
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
CC  
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@sib-sib.ch).  
CC  
CC EMBL: Z15005; CAA78727.1; -  
CC PIR: S28261; S28261.  
CC HSSP: P17119; 3KAR.  
CC MIM: 117143; -  
CC InterPro: IPR001752; kinesin.  
CC Pfam: PF00225; kinesin; 1.  
CC PRINTS: PR00380; KINESINHEAVY.  
CC SMART: SM00129; KISC; 1.  
CC PROSITE: PS00411; KINESIN_MOTOR_DOMAIN1; 1.  
CC PROSITE: PS00067; KINESIN_MOTOR_DOMAIN2; 1.  
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
KW Cell cycle; Centromere.  
FT DOMAIN 1 335 KINESIN-MOTOR.  
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).  
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).  
FT NP_BIND 86 93 ATP (BY SIMILARITY).  
SQ SEQUENCE 2663 AA; 312087 MW; CEPCL3880C8C8C8 CRC64;  
  
Query Match 5.4%; Score 142.5; DB 1; Length 2663;  
Best Local Similarity 21.5%; Pred. No. 0.56;  
Matches 111; Conservative 81; Mismatches 191; Indels 133; Gaps 23;  
  
QY 54 KDKTAISENLKSFSE-----VGEIDA-----DEEVKALTGKQMK----- 89  
DB 1200 KERVKLKELQSFETDRHLRGYIRETEATGLQKEELKIAHILKHEQETIDELRRSVS 1259  
QY 90 -----IMMERKEHEHNLSTLKKCKREKQEAAL---KLLNEVQHELEERLCRESLA- 139
```





Db 1551 KEASLSEWLSATETELVQKSTSEGLGLDLDTEISNAKNVLKDL-----KKADLNTIT 1605  
 Qy 105 TLKCKREEKEALKLNEVQEHLEERLCRESLADSWGECRSLNNCMRIYTTCPSPW 164  
 Db 1606 -----ESSAALQNLIEGSEPI--LEBRLC--VLNAGWSRVRTWTDWCNTL----- 1647  
 Qy 165 SSVKNKTERFRKTI-----YQFLPFPHEDNEKDLPISEKLTIE-----EDAQLTQMED 211  
 Db 1648 MNHQNLIEFDGNVAHISTWLYQAEALLDEIEKKPTSKQEIYKRLVSELDNDANL-QVEN 1706  
 Qy 212 VFSQTLVDVNSLNFNRNFVQMQOEDQTFQ--SHFISDTDLTEPYFFPAFSKEPMTKA 269  
 Db 1707 VRDQALILNARGSSRELVEPKLAELNRNFEKVSQHIKSAKL-----LIAQEP----- 1755  
 Qy 270 DLEQCWDIPNFQFCNFVSIVSYESVETITKMLKAI-----EDLPKQDKAPDHGGLLSKM 325  
 Db 1756 -LYQCLVTTTFETFGVPF--SDLEKLENDIENMLKFVEKHLSEDEDEKMEESAQIEEV 1812  
 Qy 326 LPQDGRGLGELDONLSRCFKFHEKCKQCAHLSEDCPDVPAHTEDELDAIRLVNVSNOQ 385  
 Db 1813 LQGEEMLHQPMDN-----KKEKIRQLL-----LHTRYNK-IAKAIPIQQRK 1855  
 Qy 386 YGQILOWTRKHELTAYLVEKMRQFGWVSELANQAPETELIFNSIQVVPRIHEGNISQ 445  
 Db 1856 MGQLASGIRSSLLPTDYLVE-----INKILLCMDVLSLNPVELNTAIYE-DFSQ 1906  
 Qy 446 DETM 449  
 Db 1907 EDSL 1910

RESULT 15  
 RA50\_METJA  
 ID RA50\_METJA STANDARD; PRT: 1005 AA.  
 AC Q58718;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE DNA double-strand break repair rad50 AtPase.  
 GN RAD50 OR MJ1322.  
 OS Methanococcus jannaschii.  
 OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
 OC Methanococcus.  
 OX NCBI\_TaxID=2190;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAL-1 / DSM 2661 / ATCC 43067;  
 RX MEDLINE=96337999; PubMed=8688087;  
 RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
 RA Sutton G.G., Blake J.A., FitzGerald L.M., Clayton R.A., Gocayne J.D.,  
 RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
 RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
 RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
 RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
 RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
 RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;  
 RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
 jannaschii";  
 RL Science 273:1058-1073(1996).  
 CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
 CC rad50/mrell complex possesses single-strand endonuclease activity  
 CC and ATP-dependent double-strand-specific exonuclease activity.  
 CC Rad50 provides an ATP-dependent control of mreII by unwinding  
 CC and/or repositioning DNA ends into the mreII active site (By  
 CC similarity).  
 CC -!- SUBUNIT: Forms a complex with mreII (By similarity).  
 CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 DR EMBL: U67572; AAB99331.1; -.  
 DR TIGR: MJ1322; -.  
 DR InterPro: IPR003439; ABC\_transportr.  
 DR InterPro: IPR001687; ATP\_GTP\_A.  
 DR InterPro: IPR003405; SMC\_C.  
 DR InterPro: IPR003395; SMC\_N.  
 DR Pfam: PF02483; SMC\_C; 1.  
 DR Pfam: PF02463; SMC\_N; 1.  
 DR DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
 KW NP\_BIND 32 39 ATP (By SIMILARITY).  
 FT DOMAIN 158 849 COILED COIL (POTENTIAL).  
 FT SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.4%; Score 141; DB 1; Length 1005;  
 Best Local Similarity 19.8%; Pred. No. 0.22;  
 Matches 90; Conservative 80; Mismatches 172; Indels 112; Gaps 17;  
 Qy 38 IVCLLLKDSHCAPTWKDKTAISENLKSFSEVGEIDAEEVKKALTGKQKIMMERKEK 97  
 Db 137 IAKFLSLKPSKLETVAKLIGIDEFEKCYQKMGEL--VKEYERLERIEGELNKENYEK 194  
 Qy 98 EHTNLMSTLKKCKREEKQEAALKLLNEVQEHLEERLCRESLADSWGECRSLNNCMRIY 157  
 Db 195 ELKNKMSQLEKNKMLMEINDKLNKIKKEFEDIEKLFNE----- 233  
 Qy 158 TTCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISE---KLIEDAQLTQMEDVFS 214  
 Db 234 -----WENKKLYEKFINL-----BERKRALKLNQELKILEYDL----- 269  
 Qy 215 QLTVDVNSLNFNRNFVQMQOEDQ-----TFQSHFISDTDLTEPYFFPAFSKEP 265  
 Db 270 NTVVEARETLNRHKDEYEKYSLVDEIRKTESRURELKSHEVEDYKLTGK-----QLE 321  
 Qy 266 MTKADLEQCWDIPNFQFCNFVSIVSYESVETITKMLKAIED-LPKQDKAPDHGGLISK 324  
 Db 322 IIKGDIKLEKFEIN-----KSKYRDDIDNLDLTLNKKIKIEVETIKD---LLEE 369  
 Qy 325 MLPQDGRGLGELDONLSRCFKFHEKCKQCAHLSEDCPDVPAHTEDELDAIRLVNVSNO 384  
 Db 370 L-----KNLNEEIEKIEKRYKRICECKEYEEK-----YLEEEKAVYENKLT 412  
 Qy 385 QYGQILOWTR---KHELTAYLVEKMRQFGWVSELANQAPETELIFNSIQV---PRIH 438  
 Db 413 EYITLLQEKKSIEKNINDLETRINKL-----LBETKN--IDIESIENSLKEIEKKVL 464  
 Qy 439 EGNISKQDETMMDLSILPSSNFTLKIPLESAAE 472  
 Db 465 E-NLQKEKIELNKKLGEINSEIKRLKKILDELKE 497

Search completed: July 2, 2002, 11:50:18  
 Job time: 377 sec

**THIS PAGE BLANK (USPTO)**

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:43:26 ; Search time 53.01 seconds  
(without alignments)  
1615.403 Million cell updates/sec

Title: US-09-722-544A-2

Perfect score: 2632

Sequence: 1 MKIKAEKNEGPRSRWQLHW.....FICYVAKALQHKHEKTKW 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*

1: sp\_archaea:\*

2: sp\_bacteria:\*

3: sp\_fungi:\*

4: sp\_human:\*

5: sp\_invertebrate:\*

6: sp\_mammal:\*

7: sp\_mhc:\*

8: sp\_organelle:\*

9: sp\_phase:\*

10: sp\_plant:\*

11: sp\_rodent:\*

12: sp\_virus:\*

13: sp\_vertebrate:\*

14: sp\_unclassified:\*

15: sp\_rvirus:\*

16: sp\_bacteriaph:\*

17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	2464	93.6	466	4	Q15846
2	1836.5	69.8	465	6	Q95KN1
3	1230.5	46.8	338	6	Q9N1T8
4	423	16.1	449	6	Q29482
5	401.5	15.3	448	13	Q9RGPO
6	321	12.2	372	11	Q9JK98
7	266.5	10.1	295	11	Q35510
8	204	7.8	218	11	Q9ERD1
9	160	6.1	1738	5	O76329
10	156.5	5.9	3616	13	Q9W6V0
11	151.5	5.8	1031	5	Q18082
12	149.5	5.7	840	6	Q95JRO
13	146	5.5	1156	16	O66878
14	145	5.5	1388	6	Q28021
15	144	5.5	1305	10	Q9FJ35
16	144	5.5	3259	4	Q14789

17	143	5.4	1931	5	Q9NCF9	Q9ncf9 drosophila
18	143	5.4	1931	5	Q9VRH9	Q9vkh9 drosophila
19	142.5	5.4	533	6	Q95JY2	Q95jy2 macaca fasc
20	141	5.4	1388	4	Q75116	Q75116 homo sapien
21	141	5.4	1388	4	Q9UON5	Q9uon5 homo sapien
22	140.5	5.3	720	4	Q9H6Q7	Q9h6q7 homo sapien
23	140.5	5.3	2166	16	O51465	O51465 borrelia bu
24	140	5.3	1379	11	O62868	O62868 rattus norv
25	139	5.3	1330	6	O97961	O97961 vulpes vulp
26	139	5.3	1530	4	O43241	O43241 homo sapien
27	139	5.3	1955	5	O61308	O61308 parascaris
28	139	5.3	5458	5	O9U459	O9u459 plasmodium
29	138.5	5.3	1391	11	O922J3	O922j3 mus musculu
30	137	5.2	1084	16	O83423	O83423 treponema p
31	137	5.2	1300	4	Q13999	Q13999 homo sapien
32	136.5	5.2	709	5	O9GRG1	O9grg1 tetrahymena
33	136	5.2	1132	4	O75065	O75065 homo sapien
34	135.5	5.1	470	10	Q38843	Q38843 arabidopsis
35	135.5	5.1	684	3	Q07238	Q07238 pneumocysti
36	134.5	5.1	882	17	Q96YR5	Q96yr5 sulfolobus
37	134.3	5.1	1708	5	Q9U0S6	Q9u0s6 mytilus gal
38	134.5	5.1	1723	2	Q9JMX8	Q9jmx8 helicobacte
39	134.5	5.1	1958	5	O96062	O96062 dugesia jap
40	134	5.1	1356	4	Q14707	Q14707 homo sapien
41	133.5	5.1	1956	5	Q20641	Q20641 caenorhabdi
42	132.5	5.0	1410	4	Q14221	Q14221 homo sapien
43	132	5.0	1033	4	Q9UFE1	Q9ufe1 homo sapien
44	132	5.0	1133	5	Q21022	Q21022 caenorhabdi
45	132	5.0	1270	4	Q96JN2	Q96jn2 homo sapien

#### ALIGNMENTS

RESULT 1

Q15846 PRELIMINARY; PRT; 466 AA.  
AC Q15846;  
DT 01-NOV-1996 (TRENBLrel. 01, Created)  
DT 01-NOV-1996 (TRENBLrel. 01, Last sequence update)  
DT 01-DEC-2001 (TRENBLrel. 19, Last annotation update)  
DE HYPOTHETICAL 54.2 KDA PROTEIN PRECURSOR.  
OS Homo sapiens (Human).  
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
OX NCBI\_TaxID=9606;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=NEURORETINA;  
RA Shimizu A., Nishida K., Kinoshita S., Inazawa J., Okubo K.,  
RA Matsubara K.;  
RT "Expression profile of active genes in human retina.";  
RL Submitted (AUG-1995) to the EMBL/GenBank/DBJ databases.  
DR EMBL; D63813; BAA09882.1; -;  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 3.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
KW Signal; Hypothetical protein. POTENTIAL.  
FT SIGNAL 1 20  
FT CHAIN 21 466 ROD PHOTORECEPTOR PROTEIN.  
SQ SEQUENCE 466 AA; 54214 MW; 41A603D0FD6439C4 CRC64;

Query Match 93.6%; Score 2464; DB 4; Length 466;  
Best Local Similarity 100.0%; Pred. No. 2.3e-161;  
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 30 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENLSKFSFSEVGEIDADEEYKALTGKQMK 89

Db 1 MKPPLLVFVCLLWLKDSHCAPTWKDKTALSENLSKFSFSEVGEIDADEEYKALTGKQMK 60

QY 90 IMMERKEKHTNLMSTLKKCRREEKQKALLLNEVQVHLEERLCRESLADSNNGCRSCL 149

```
Db 61 IMMEREKEHTNLMSTLKKREKQKALKLLNEVOEHLEEEERLCRESLADSWGCRSCL 120
QY 150 ENNCMRIYTTCPSSVSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOM 209
Db 121 ENNCMRIYTTCPSSVSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOM 180
QY 210 EDVFSQTLVDVNSLNRSNFVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269
Db 181 EDVFSQTLVDVNSLNRSNFVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 240
QY 270 DLEQCHWDINPQLFCNFSVSYESVETITKMLKAIEDLPKQDAPDHGGLISKMLPGQ 329
Db 241 DLEQCHWDINPQLFCNFSVSYESVETITKMLKAIEDLPKQDAPDHGGLISKMLPGQ 300
QY 330 DRGLCGELQDLNLSRCFKFEKQKQCAHLSDECDPVPALHTLDELDAIRLVNVSNOOYGOI 389
Db 301 DRGLCGELQDLNLSRCFKFEKQKQCAHLSDECDPVPALHTLDELDAIRLVNVSNOOYGOI 360
QY 390 LQMTKRKHEDTAYLVKEMRGQFGWSELANQAPETEIIFNSIQVVPRIHEGNISKODETM 449
Db 361 LQMTKRKHEDTAYLVKEMRGQFGWSELANQAPETEIIFNSIQVVPRIHEGNISKODETM 420
QY 450 MTDLSILPSSNFTLKIPLEESAESSNFICYVYVAKALQHKEHFKTW 495
Db 421 MTDLSILPSSNFTLKIPLEESAESSNFICYVYVAKALQHKEHFKTW 466

RESULT 2
Q95KNI PRELIMINARY; PRT; 465 AA.
ID Q95KNI
AC Q95KNI;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE RETINAL CLUSTERIN-LIKE PROTEIN CLUL1B SPLICE VARIANT.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=RETINA.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF241221; AAK49030.1; -.
DR SMART; SMART; SM00030; CLB; 1.
SQ SEQUENCE 465 AA; 54402 MW; BBDE1AC512D5D33F CRC64;

Query Match 69.8%; Score 1836.5; DB 6; Length 465;
Best Local Similarity 73.4%; Pred. No. 2.7e-118;
Matches 342; Conservative 51; Mismatches 72; Indels 1; Gaps 1;

QY 30 MKPPLLVFIVCLWLKDSHCAPTWKDKTALSENKLSFSEVGEADAEVVKALTGKQMK 89
Db 1 MKPPLLVFIVCLWLKDSHCAPTWKDKTALSENKLSFSEVGEADAEVVKALTGKQMK 60
QY 90 IMMEREKEHTNLMSTLKKREKQKALKLLNEVOEHLEEEERLCRESLADSWGCRSCL 149
Db 61 IMMEREKEHTNLMSTLKKREKQKALKLLNEVOEHLEEEERLCRESLADSWGCRSCL 120
QY 150 ENNCMRIYTTCPSSVSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOM 209
Db 121 ENNCMRIYTTCPSSVSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOM 180
QY 210 EDVFSQTLVDVNSLNRSNFVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269
Db 181 ENVFNQTLVDVNSLNRSNFVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKA 240
QY 270 DLEQCHWDINPQLFCNFSVSYESVETITKMLKAIEDLPKQDAPDHGGLISKMLPGQ 329
```

```
Db 241 DPVQSWDIPSPFQFLPYNFSLSIYHSITITTKLNAIEDLPKQDNDSNHGSLSSKTLPVQ 300
QY 330 DRGLCGELQDLNLSRCFKFEKQKQCAHLSDECDPVPALHTLDELDAIRLVNVSNOOYGOI 389
Db 301 HRPVGEFGONLSECFQFHARCKQKODYLVWEDCPDVPALHTKVDEALVNLVSHQOYQV 360
QY 390 LQMTKRKHEDTAYLVKEMRGQFGWSELANQAPETEIIFNSIQVVPRIHEGNISKODETM 449
Db 361 LQMTQHLEDITLYLMEKMRREEFGWADLANQAPGAENIFDSTKMPVNIHEGNFSKODETM 420
QY 450 MTDLSILPSSNFTLKIPLEESAESSNFICYVYVAKALQHKEHFKTW 495
Db 421 I-DLSILSPSPNFTLKIPLEESAETSNFISYMLEKAVQHFKHFKTW 465

RESULT 3
Q9NIT8 PRELIMINARY; PRT; 338 AA.
ID Q9NIT8
AC Q9NIT8;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE RETINAL-SPECIFIC CLUSTERIN-LIKE PREPROTEIN.
GN CLUL1.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Cranialata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=20156379; PubMed=10675623;
RA Zhang Q., Ray K., Acland G.M., Czarnecki J.M., Aguirre G.D.;
RT "Molecular cloning, characterization and expression of a novel retinal
clusterin-like protein cDNA.";
RL Gene 243:151-160(2000).
DR EMBL; AF147784; AAF36799.1; -.
DR InterPro; IPR000753; Clusterin.
DR SMART; SM00035; CLA; 1.
DR SMART; SM00030; CLB; 1.
SQ SEQUENCE 338 AA; 39385 MW; B86F07877213FCC2 CRC64;

Query Match 46.8%; Score 1230.5; DB 6; Length 338;
Best Local Similarity 70.1%; Pred. No. 7.3e-77;
Matches 232; Conservative 41; Mismatches 57; Indels 1; Gaps 1;

QY 165 SSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOMEDVFSQTLVDVNSLF 224
Db 9 TSGTSTVEQFFRNIIYQFLFPFHEDNEKDLPISEKLIBEDAQLTOMEDVFSQTLVDVNSLF 68
QY 225 NRSFNVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEQCNWDIPNFQFL 284
Db 69 NRSFNVRMQQEFQDTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEQCNWDIPNFQFL 128
QY 285 CNFVSIIYSVSETITKMLKAIEDLPKQDAPDHGGLISKMLPGCDGLCGELQDLNLSRC 344
Db 129 YNFSLSIYHSITITTKLNAIEDLPKQDNDSNHGSLSSKTLDPVQHRGYPGEFGGLSEC 188
QY 345 FPFHEKQKQCAHLSDECDPVPALHTLDELDAIRLVNVSNOOYQQLQMTKRKHEDTAYLV 404
Db 189 FQFHARCKQKODYLVWEDCPDVPALHTKVDEALVNLVSHQOYQVLMQTHLEDITLYM 248
QY 405 EKMGRGFGWSELANQAPETEIIFNSIQVVPRIHEGNISKODETMVTDLSILPSSNFTLK 464
Db 249 EKMREEFGWADLANQAPGAENIFDSTKMPVNIHEGNFSKODETMI-DLSILSPSPNFTLK 307
QY 465 IPLEESAESSNFICYVYVAKALQHKEHFKTW 495
Db 308 IPLEESAESSNFICYVYVAKALQHKEHFKTW 338

RESULT 4
```

```

Q29482
ID Q29482 PRELIMINARY; PRT; 449 AA.
AC Q29482;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE CLUSTERIN PRECURSOR.
OS Equus caballus (Horse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Perissodactyla; Equidae; Equus.
OX NCBI_TaxID=9796;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-OLIVER; TISSUE-TESTIS;
RA Barber J.A., Farris J.A., Troedsson M.H.T., Crabo B.G., Foster D.;
RT "Nucleotide sequence of the complementary DNA encoding Equine
  Clusterin.";
RL Submitted (NOV-1995) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
  SIMILARITY).
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC EMBL: L46797; AAA80313.1; -
DR InterPro: IPR000753; Clusterin.
DR Pfam: PF01093; Clusterin; 1.
DR SMART: SM00035; Cla; 1.
DR SMART: SM00030; CLB; 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Signal
FT SIGNAL 1 21 POTENTIAL.
FT CHAIN 22 449 CLUSTERIN.
SQ SEQUENCE 449 AA; 52154 MW; 705A60504515F7C0 CRC64;

```

Query Match 16.1%; Score 423; DB 6; Length 449;  
Best Local Similarity 24.5%; Pred. NO. 2.7e-21;  
Matches 117; Conservative 100; Mismatches 203; Indels 58; Gaps 13;

QY	34	LLVFIYVCLLWLKDSHCAPTWKOKTALISENLKSFSEVGEIDAEDEEVKALTGIKOKIMME	93
Db	4	LLLVGLLLTLENGQ---VLGDKAVSDRELQEMSTQGSNYINKETNALKVGCKQIKNLE	60
QY	94	RKEKEHTNLWSTLUKKCREKQEQALKLLNEVOHLEBEERLCRESLADSWGECRSCLENNC	153
Db	61	QTNEERKSLLGTLEEAKKKEGALNDTKOSEMKLKESQVCNETHALWEECPCLKQTC	120
QY	154	MRIYT--TCOPSSWVKNKIETFRFKTIYQFLFPFHEDNEKDLPISEKLIEDAQLTQMEDV	212
Db	121	MKEFYARVCRSGSLGVGHQLEELFNQSSPFYFWINGDR-----IDSLLENDRQOQTHVLVD	174
QY	213	FSOLTVDVNSLFRNSFNVPFRMQQEPDFONFOSHETSDTLTEPYEPFAPSKPEPMFKADL-	271
Db	175	-----MQDSFDRASSI-----MDELUFQDRFTT-REPQDTYIYSPPS-SPHRRSLL	218
QY	272	-----BOCWDP------NFFQLFCNFSVIYESVSETITMKLKAIEDLPKQD---RAP	316
Db	219	FNPKSRFARNIMHFFMYRHLNFDMFOPEFDIMHOAAQAMNLLHRLPDQLPMTFESEGD	278
QY	317	DHGGLISKMLPGDGRGLCGELONLSRCFKFEHKCKQCAHLSEDC---PDVPALHTL	372
Db	279	NH-----DRTVCKEIRHNSGCGLUKMDQCEKQCEILSVDCSTNNPSQMQLRQEL	327
QY	373	DEAIRLVNSNQOYGIOQLMTRKRLHEDTAYLVLEKMRQFGWYSELANAPEETIFNSIQ	432
Db	328	NNSLQAEKFTKLYDELLQSYQEKMLNTSLLKQLNEQFSWYSQLANLTQGEDOYYLQVT	387
QY	433	VVPRIHEGNISKODETMMTDLSLTLPSSNTLKIPIEESAESSNFTGYVYVAKALQHFKE	490
Db	388	TVSS--INSDOSEVPSGLTRVVVKLFDSPYITVATVPEVVSNNKPFMETVAEKALQYRV	444

RESULT	5	
Q9YGP0		
ID	Q9YGP0	PRELIMINARY; PRT; 448 AA.
AC	Q9YGP0;	
DT	01-MAY-1999	(TReMBLrel. 10, Created)
DT	01-MAY-1999	(TReMBLrel. 10, Last sequence update)
DT	01-DEC-2001	(TReMBLrel. 19, Last annotation update)
DE	CLUSTERIN.	
OS	Gallus gallus	(Chicken).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;	
OC	Gallus.	
OX	NCBI_taxID=9031;	
RN	[1]	
RP	SEQUENCE FROM N.A.	
RX	MEDLINE=99134326; PubMed=9933595;	
RA	Mahon M.G., Lindstedt K.A., Hermann M., Nimpf J., Schneider W.J.;	
RT	*Multiple involvement of clusterin in chicken ovarian follicle	
RT	development. Binding to two oocyte-specific members of the low density	
RT	lipoprotein receptor gene family.*;	
RL	J. Biol. Chem. 274:4036-4044(1999).	
CC	-1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY	
CC	OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,	
CC	AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED	
CC	CELL DEATH (BY SIMILARITY).	
CC	-1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY	
CC	SIMILARITY).	
CC	-1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.	
DR	EMBL: AF119370; RAD17257.1; ..	
DR	InterPro: IPR000753; Clusterin.	
DR	Pfam: PF01093; Clusterin; 1.	
DR	SMART: SM00035; CLA; 1.	
DR	SMART: SM00030; CLB; 1.	
DR	PROSITE: PS00492; CLUSTERIN_1; 1.	
DR	PROSITE: PS00493; CLUSTERIN_2; 1.	
KW	Glycoprotein.	
SQ	SEQUENCE 448 AA; 51348 MW; 9982873DCA13D27C CRC64;	

Query Match	15.3%;	Score	401.5;	DB	13;	Length	448;
Best Local Similarity	26.2%;	Pred.	No. 8.1e-20;				
Matches	121;	Conservative	86;	Mismatches	178;	Indels	77;
Gaps	17;						
QY	63	LKSFSEGEIDADREVKALGIQKIMTKMERKKEHTNLMSTLKKCKREEQKALKLINE	122				
DB	27	LKQLSAGSKVIDEVENAINGVOMKTLMDKTSKEHQAMLHTLEETKRRKEEAVKALE	86				
QY	123	VQEHLEERLCRSLADSWGECRSLENNCMRIYT-TCOPSWSVKNKIBFRFKIYQF	181				
DB	87	KEKOLAEKQECVNTMLSLWPECKPLKHTCMRVYSKICHSGSLGVGRQLEELLNRSPPF	146				
QY	182	-----LFFFHEDNEK---DLPISEKLIREDQALQOMEDVFSOLTVDVNSLNFNR	227				
DB	147	SIWVNGREIDALLDREQQRERFDEERFGIMEDG-----VEDIFQ-----DSTQLYGA	197				
QY	228	FNVFRMQQEPDQTFQSHFISDTDLTEPY---FFFAFSKEPMTKADLEQCWDIPNFFQL--	283				
DB	198	FPFPR-----TTPFGGFEAF-VPPVQVRVL-----VPPRRLSR	231				
QY	284	----FCNFSYIYESVSETITKMLKALIEDLPQDKADPH---GGLISKMLP-GQDRGLCGE	336				
DB	232	ELHPFLQHPVGHFRLFEMTQRMIDG-----GHGAMDHLGLGFESESRNFSTDRVMGRC	285				
QY	337	LDQNLSCRFHFHEKQCAHLSDC----PDVPAALITELDEARLVNVSNOOQYQIQLQ	392				
DB	286	IRRSAGCLMRDECKEREILAVDCSQTDVPQSLQREQFEDALRLAERFTRYDDLLSA	345				
QY	393	TRKHELDATVLEKMRQCFGWSELANQAPTEIIFNSIQV---VPRITHEGNISKQDET	449				
DB	346	QFAEMLNTSLLDQNLRFQGVRLGNLTQCTDGLQVTTVFSTKPTNLEDP--SAPADTQ	403				

QY 450 MTDLSILPSSNFTLKIPLSEASSNFICYVYAKALQHFKEH 491  
DB 404 VT-VQLFDSEPLSLVPGDISWDDPRFMEIVAQAALQHYKON 444

## RESULT 6

Q9JK98 PRELIMINARY; PRT; 372 AA.

AC Q9JK98;  
DT 01-OCT-2000 (TREMBlrel. 15, Created)  
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)  
DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)  
DE CLUSTERIN ISOFORM 2.

OS Mus musculus (Mouse).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI\_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=BALB/C; TISSUE=UTERUS;

RA You K.H.; Jeon J.H.;

RT "Identification of Truncated SGP-2 Lacking a Signal Peptide for

RT Secretion.";

RL Submitted (MAR-2000) to the EMBL/GenBank/DDAJ databases.

CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH (BY SIMILARITY).

CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY

CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

DR EMBL: AF248058; AAF67185.1; -.

DR InterPro: IPR000753; Clusterin.

DR Pfam: PF01093; Clusterin; 1.

DR SMART: SM00035; Cla; 1.

DR SMART: SM00030; Cla; 1.

DR PROSITE: PS00492; CLUSTERIN\_1; 1.

DR PROSITE: PS00493; CLUSTERIN\_2; 1.

KW Glycoprotein.

SQ SEQUENCE 372 AA; 43109 MW; A104A7E48520FEFE CRC64;

Query Match 12.2%; Score 321; DB 11; Length 372;

Best Local Similarity 23.6%; Pred. No. 2.2e-14;

Matches 93; Conservative 83; Mismatches 170; Indels 48; Gaps 12;

QY 113 KQALKLLNVEQHEERLCRESLADSWGECRCLNENCMRIYT-TCQPSWSSVKNKI 171

DB 4 REDALEDTRESEMKLKAFPEVCNETMMALWEECKPCLKHTCMFYARVCRSGSLVGQQL 63

QY 172 ERFRKIYQFLPFHEDNEKDLPISEKLIBEADAQLTQ---MEDVFSQLTVDVNSLFNRS 227

DB 64 EEFNLSPPFFWMNGDR-----IDSLESDRQSQVLDAMQDSFARASGIIDTLFQDR 117

QY 228 ENVFROMQOEFDOTFQSHFTSDTLTEPYFPF-----AFSKEPMTKADLEQC-WDIPNF 280

DB 118 F--FAR-----ELHDPHYFS-----PIGFPHKRPHFYLPKSLRSLMSPHYGPSPF 163

QY 281 FQLFCNFSVIYSVSSTITKMLKAIEDLPKQKADPHGGLISKMLPGQ-DRGLCGELDQ 339

DB 164 HNMFPFFMIHQAAQAMDVLHSPAQFPDQD-----FLREGEDDRTVCKEIRR 213

QY 340 NLSCFKFHEKCKQCAHLSDEC-----PDVPAHLTELDEAIRLVNSNQYGOILQWTRK 395

DB 214 NSTGCLMKKGQCEKQCEILSDVDCSTNNPAQANLRQELNDSLQVAERTEQYKELLSFQS 273

QY 396 HLEDATYLVKMGQFQSWSELANAQAPETIIFNSIQVPRIRHEGNISKODEMTMTDLSI 455

DB 274 KMLTSSLLQLNDQFNWSQLANLQGEDKYVLRVSTV-TTHSSDSEVPSTVEVVVKL 332

QY 456 LPSNFTLKIPLSEASSNFICYVYAKALQHFKEH 489

DB 333 FDSDPITVWLPEEVSQKPNKFMDTVAEKALQEVYR 366

## RESULT 7

O35510

AC O35510 PRELIMINARY; PRT; 295 AA.

DT 01-JAN-1998 (TREMBlrel. 05, Created)

DT 01-JAN-1998 (TREMBlrel. 05, Last sequence update)

DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)

DE CLUSTERIN.

OS Rattus norvegicus (Rat).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.

OX NCBI\_TaxID=10116;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=WISTAR; TISSUE=SEMINAL VESICLE;

RX MEDLINE=98163873; PubMed=9503143;

RA Izawa M.;

RT "Identification of a transcript predicting an alternative form of

RT sulfated glycoprotein-2 (clusterin) in rat tissues.";

RL Biochem. Mol. Biol. Int. 44:9-19(1998).

CC -1- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY

OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,

AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED

CELL DEATH (BY SIMILARITY).

CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY

CC SIMILARITY).

CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

DR EMBL: D11339; BAA21775.1; -.

DR InterPro: IPR000753; Clusterin.

DR Pfam: PF01093; Clusterin; 1.

DR SMART: SM00035; Cla; 1.

DR SMART: SM00030; Cla; 1.

DR PROSITE: PS00492; CLUSTERIN\_1; 1.

DR PROSITE: PS00493; CLUSTERIN\_2; 1.

KW Glycoprotein.

SQ SEQUENCE 295 AA; 34115 MW; C1E07FA745DD5470 CRC64;

Query Match 10.1%; Score 266.5; DB 11; Length 295;

Best Local Similarity 25.2%; Pred. No. 9.1e-11;

Matches 80; Conservative 59; Mismatches 131; Indels 47; Gaps 11;

QY 133 LCRESLADSWGECRCLNENCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPFHEDNEK 191

DB 9 VCNETMMALWEECKPCLKHTCMKIFYARVCRSGSLVGROLEEFLNQSPPFYFMMNGDR-- 66

QY 192 DLPISKEKLEEDAQLTQ---MEDVFSQLTVDVNSLFNRSFNVFROMQOEFDQTFQSHFI 247

DB 67 ----IDSLESDRQSQVLDAMQDSFTRASGIIDTLFQDRF--FTHEPQDI-----HHF- 114

QY 248 SDTDLTEPYFPF-----AFSKEPMTKADLEQCWDIP-NFFQLFCNFSVIYSVSSETIT 300

DB 115 -----SPMGFPHKRPHFYLPKSLRSLRSLMPLSHYGLPSFHNMFQPFDFMIHQAAQAMD 168

QY 301 KMLKATEDLPKQKADPHGGLISKMLPGQ-DRGLCGELDONLSRCFKFHEKCKQCAHLS 359

DB 169 QLHSPALQFPDQVDFLKE-----GEDDPTVCKEIRHNSTGCLMKMGQCEKCEILS 218

QY 360 EDC-----PDVPAHLTELDEAIRLVNSNQYGOILQWTRKHLDEDTAYLVKMKMGQFGWYS 415

DB 219 VDCSTNNPAQANLRQELNDSLQVAERLTQQYNELLHSLOSKMLNTSSLLQLNDQFSWVS 278

QY 416 ELANQAPETIIFNSIQ 432

DB 279 QLANLTQGGDDQYFGSPQ 295

## RESULT 8

Q9ERD1

ID Q9ERD1 PRELIMINARY; PRT; 218 AA.

AC Q9ERD1;

DT 01-MAR-2001 (TrEMBLrel. 16, Created)  
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE CLUSTERIN (FRAGMENT)  
OS Rattus norvegicus (Rat).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.  
OX NCBI\_TaxID=10116;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN-SPRAGUE DAWLEY; TISSUE=THYMUS;  
RA Park J.H., Park J.S., Ju S.K., Na S.Y., You K.H.;  
RT "Determination of clusterin mRNA expression of apoptosis induced rat  
thymocytes in vivo and in vitro."  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CELL DEATH (BY SIMILARITY).  
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY  
SIMILARITY).  
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
DR EMBL; AF314657; AAG31162.1; -.  
DR InterPro; IPR00753; Clusterin.  
DR Pfam; PF01093; Clusterin; 2.  
DR SMART; SM00035; CLA; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Glycoprotein.  
FT NON\_TER 1 1  
SQ SEQUENCE 218 AA; 24931 MW; ED2FC4425A510589 CRC64;

Query Match 7.8%; Score 204; DB 11; Length 218;  
Best Local Similarity 28.3%; Pred. No. 1.2e-06;  
Matches 47; Conservative 36; Mismatches 73; Indels 10; Gaps 3;

QY 330 DRGCGELDQNLSCRFHEKCKQAHLSDC-----PDVPALHTEDEAIRLVNVNQ 385  
DB 51 DRTCKEIRHNSGCLKMKGCKCEKCEILSVDCSTNPAQANLQSLQVAERLQ 110  
QY 386 YGOIQMTRKHLDTAVLVEKMRQFGWVSELAN--QAPETIIFNSQVVPRIHEGNIS 443  
DB 111 YNELLHSLSQKMLNTSLLEQLNDFSWVSQLANTQDDQYLRVSIVTT-----HSSDSE 166  
QY 444 QDETMTDLSILPSSNFTLKIPLEESAENSFTGYVVAQALQHF 489  
DB 167 VPSRVTEVVKLFDSDPITVVLPEVSKDNPKFMDTVAERALQERY 212

RESULT 9  
O76329 PRELIMINARY; PRT; 1738 AA.  
ID O76329  
AC O76329  
DT 01-NOV-1998 (TrEMBLrel. 08, Created)  
DT 01-NOV-1998 (TrEMBLrel. 08, Last sequence update)  
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE INTERAPTIN.  
GN ABPD.  
OS Dictyostellium discoideum (Slime mold).  
OC Eukaryota; Mycetozoa; Dictyosteliida; Dictyostellium.  
OX NCBI\_TaxID=44689;  
RN [1]  
RP SEQUENCE FROM N.A.  
RA MEDLINE=98365466; PubMed=9700162;  
RX Rivero F.J., Kuspa A., Brokamp R., Matzner M., Noegel A.A.;  
RT "Interaptin, an actin-binding protein of the alpha-actinin superfamily  
in Dictyostellium discoideum, is developmentally and CAMP-regulated and  
associates with intracellular membrane compartments.";  
RL J. Cell Biol. 142:735-750(1998).  
DR EMBL; AF057019; AAC34582.1; -.  
DR HSSP; Q01082; IBKR.  
DR InterPro; IPR001589; Actinin\_act\_bind.  
DR InterPro; IPR001715; Calponin\_hom.

DR InterPro; IPR001990; Granin.  
DR InterPro; IPR001451; Hexapep\_transf.  
DR InterPro; IPR002017; Spectrin.  
DR Pfam; PF00307; CH; 2.  
DR SMART; SM00033; CH; 2.  
DR PROSITE; PS00019; ACTININ\_1; UNKNOWN\_1.  
DR PROSITE; PS00020; ACTININ\_2; 1.  
DR PROSITE; PS00021; CH; 2.  
DR PROSITE; PS00422; GRANINS\_1; UNKNOWN\_1.  
DR PROSITE; PS00101; HEXAPEP\_TRANSFERASES; UNKNOWN\_1.  
SQ SEQUENCE 1738 AA; 204427 MW; 577A99D2EC79AF5C CRC64;

Query Match 6.1%; Score 160; DB 5; Length 1738;  
Best Local Similarity 20.9%; Pred. No. 0.016;  
Matches 102; Conservative 88; Mismatches 171; Indels 128; Gaps 21;

QY 54 KDRTAISENL-----KSFSEVGEIDADEVKALTIKIGIKOMKIMMER----- 94  
DB 667 KDNQTINEQLKQLSEKDEKIEKLSNQEQOODEKINNLLLEKEKCLIERINQQLLEN 726  
QY 95 -----KEKHTNLMSTLKKREEKQKALKLLNEVQEHLEEEER 132  
DB 727 IDLSKYQQLLLEFENFKLSSKE-----NQLNELQSKQDERFNQLND--EKLEKEKQ 779  
QY 133 LCRESLADSWGECRSCLENNCMRIYTTTCQPSWSSVKNKIERFFRKIYQFLFPFHEDNEKD 192  
DB 780 L--QSIEDFNQYK-----QQQLSSNSN-IDQQLQSTIIESELKEQKELN 822  
QY 193 LPISEKLIIEEDAQLQTMEDVFSQLT-----VDVNSLFNRSFNVFRMQMQEFDQTFQSH 245  
DB 823 ---DSKLTIEKEKQLQQLQEQFDFDLNEKNKHQDQLELEKQ---LKLQEQEYDQLNETN 876  
QY 246 FISDTDLTEYFFPAFSKEPMTADLE-----QCWDIPNFFOLFQNFVSIVYESV 295  
DB 877 QSIENQLNQNL---INKENLKEKEQLKLQNLQNOQIEKIQFDQEQFSKQNSINIELV 933  
QY 296 SETITMKLKAIED---LPKQDKA---PDHGLITKMLPGQDGLCGCLDNLSCRPFHE 349  
DB 934 NERNEKLIQLQDYDQLKQNRNDEKENDLEK---ENQLKSIONELNOLIEKNESDHK 991  
QY 350 KQCKQAHLSDECDP-----VPALHTEDE-----AIRLVNVSNQYQGIL-----Q 391  
DB 992 EQQLKQOSIENDLIEKENIQQLQSQLNEQROQOSNQLSEKQQLNOLIEKNQFDQKEQ 1051  
QY 392 MTRKHLEDATY---LVEKMRGFGF-WVSELANQAPETIIFNSQVVPRIHEGNISKQD 446  
DB 1052 LKQOSIENDLFEKENIQQLQSQLNEQROQOSNQLSEKQ-----QQNLQILIEKNESDQK 1105  
QY 447 ETMTDLSI 455  
DB 1106 EQQLKQOSI 1114

RESULT 10  
Q9W6V0 PRELIMINARY; PRT; 3616 AA.  
ID Q9W6V0  
AC Q9W6V0  
DT 01-NOV-1999 (TrEMBLrel. 12, Created)  
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)  
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)  
DE HYPERION PROTEIN, 419 KDA ISOFORM.  
GN HYPERION.  
OS Gallus gallus (Chicken).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;  
OC Gallus.  
OX NCBI\_TaxID=9031;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC TISSUE=RETINA;  
RA Kemmer W.A., Schwarz U.;  
RT "Characterization of Hyperion, a gene coding for an abundance of gene





```

RESULT 14
Q28021 Q28021 PRELIMINARY; PRT; 1388 AA.
AC Q28021;
DT 01-NOV-1996 (TEMBLrel. 01, Created)
DT 01-NOV-1996 (TEMBLrel. 01, Last sequence update)
DT 01-DEC-2001 (TEMBLrel. 19, Last annotation update)
DE RHO-ASSOCIATED KINASE.
DS Bos taurus (Bovine).
OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Bovinae; Bos.
OX NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=96208507; PubMed=8641286;
RA Matsui T., Amano M., Yamamoto T., Chihara K., Nakafuku M., Ito M.,
RA Nakano T., Okawa K., Iwamoto A., Kaibuchi K.;
RT "Rho-associated kinase, a novel serine/threonine kinase, as a putative
RT target for small GTP binding protein Rho."
RL EMBO J. 15:2208-2216(1996).
CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES.
DR EMBL; U36909; AAC48567.1; -.
DR HSPG; Q63450; IA06.
DR InterPro; IPR002219; DAG_PE-bind.
DR InterPro; IPR000719; Euk_pkinase.

```

DR InterPro: IPR001849; PH.  
DR InterPro: IPR000961; Pkinase\_C.  
DR InterPro: IPR000861; REM\_repeat.  
DR InterPro: IPR002290; Sex\_thr\_pkinase.  
DR Pfam: PF00130; DAG\_PE\_bind; 1.  
DR Pfam: PF02185; HRI; 1.  
DR Pfam: PF00169; PH; 1.  
DR Pfam: PF00069; pkinase; 1.  
DR SMART: SM00109; CI; 1.  
DR SMART: SM00074; HRI; 1.  
DR SMART: SM00233; PH; 1.  
DR SMART: SM00220; S\_TKC; 1.  
DR SMART: SM00133; S\_TK\_X; 1.  
DR PROSITE: PSS0081; DAG\_PE\_BIND\_DOM\_2; 1.  
DR PROSITE: PSS0003; PH\_DOMAIN; 1.  
DR PROSITE: PSS00107; PROTEIN\_KINASE\_ATP; 1.  
DR PROSITE: PSS0011; PROTEIN\_KINASE\_DOM; 1.  
DR PROSITE: PSS00108; PROTEIN\_KINASE\_ST; 1.  
KW ATP-binding; Kinase; Serine/threonine-protein kinase; Transferase.  
SQ SEQUENCE 1388 AA; 160799 MW; CA83CE7D3860465D CRC64;

Query Match 5.5%; Score 145; DB 6; Length 1388;  
Best Local Similarity 16.5%; Pred.No. 0.13;  
Matches 94; Conservative 97; Mismatches 166; Indels 212; Gaps 20;

QY 54 KKTALSENKLSFSEYGEIDADEVKALTIGIKOMKI-----MMERKEKEHTNLS 104  
DB 780 KQKVLNEDVRNLT-----LKIEQETQKRLTQNDLKMQTVNTLWSEKQLKQENHLL 835  
QY 105 TLK-----KCREKQKALKLLNEVQHLSEEE-----RLCR 135  
DB 836 EMKMSLEKQNAELURKRDQADQOMKELQOLEAEQVSTLYKTQVRELKECEETKLCK 895  
QY 136 E-----SLADSMGECRSLNENCMRIYTTCPSPSSVKNKIERFFPKIYQFLFPFHEDN 189  
DB 896 ELQKKQELQDE-----RSLAAQLTILTKADSEOLARSTAEQYSDLEKE 942  
QY 190 E-KDLPIS-----KLIEDAQITQMEDVFSQLTVDVNSLFRNFSNVFROMQEFQ 240  
DB 943 KTKKELEIKEMARHKQELTEKDATIASLEETNRTLTSVDANLANEKEELNNKLKAEQ 1002  
QY 241 TTQSHFISDTLTPYFFFAFSEKPTKADLEQCNDIPNFFQLFCNFSVIVESYSETIT 300  
DB 1003 LSR---LKDEEISAAIKAFKQLTERTLK-----TQAVN 1036  
QY 301 KMLKATE-----DLPKQDAPDHGGLISKMLPGQDRLGCEL-----DQNLSCRF 345  
DB 1037 KLAIEINNRKEPVKRGNDTVRRKEK-----ENRKLHMLKSEREKLTTQMI 1082  
QY 346 KFHEKCKQCAHLSEDCPDVPAHLTEDEAIRLVNSVNOQYQILQMTKRKHE----- 398  
DB 1083 KYQKELNEMQAIAEB-----SQIRIELQTLDSKDSIDIELRSQALHGLDSSISGS 1138  
QY 399 -----DTAYLVKMRG-----QFGWVSEL----- 417  
DB 1139 PGDTADDGFPESRLEGWLSLPVRNNTKFGWKVYIVSSKKILFYDSEQDKEQSNPYM 1198  
QY 418 -----ANOAPETEIFNSIQVVPRI-----HEGNISKQDETMTDLSILPSSNFT 462  
DB 1199 VLDIDKLFHVRPVTQDVYRADAKETPRIFQILYANEGESKKEQ----- 1243  
QY 463 LKIPLEESAESSNFI---GYVAKALQHF 488  
DB 1244 ---FPVEPVGEKSNYICHKGHEFIPTLYHF 1270

RESULT 15  
Q9FJ35  
ID Q9FJ35 PRELIMINARY; PRT: 1305 AA.  
AC Q9FJ35;  
DT 01-MAR-2001 (Tremblrel. 16, Created)  
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)

DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE MYOSIN HEAVY CHAIN-LIKE PROTEIN  
OS Arabidopsis thaliana (Mouse-ear cress).  
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
OC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;  
OC eurosids II; Brassicales; Brassicaceae; Arabidopsids.  
OX NCBI\_TaxID=3702;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=COLUMBIA;  
RX MEDLINE=99087489; PubMed=9872454;  
RA Nakamura Y., Sato S., Asamizu E., Kaneko T., Kotani H., Miyajima N.,  
RA Tabata S.;  
RT "Structural analysis of Arabidopsis thaliana chromosome 5. VII.  
RT Sequence features of the regions of 1,013,767 bp covered by sixteen  
RT Physically assigned P1 and TAC clones."  
RL DNA Res. 5:297-308(1998).  
DR EMBL; AB016871; BAB10654.1; .  
SQ SEQUENCE 1305 AA; 149961 MW; F37284B8E7BFF217 CRC64;

Query Match 5.5%; Score 144; DB 10; Length 1305;  
Best Local Similarity 18.8%; Pred.No. 0.14;  
Matches 94; Conservative 82; Mismatches 176; Indels 148; Gaps 18;

QY 44 LKDSHCAPTWKOKTAISENLKLSFSEYGEIDADEVKALTIGIKOMKIMERKEKEHTNLM 103  
DB 127 LKESHSV---KERELFS---LRDIHEIHQRDSS-----TRASELAQLESKQVSDLS 174  
QY 104 STLKCRREKQKALKLLNEVQHLSEEEERLCRESLA-----DSMGECRSLNENCMRIY 157  
DB 175 ASLKAEEENKATISSKNVETMKNLEQTQITQELMAELGKLKDSHREKSELSS----- 228  
QY 158 TTCQPSWSSVKNKIERFFRKIYQFLFPFHEDNEKOLPISEKLEED-----AQLTOM 209  
DB 229 -----LVEVETHQRDSSIIHVKELEQVSSKKLVLAELNQT 264  
QY 210 -----EDVFSQLTVDVNSLFRNFSNVFROMQEFQDTFQSHFISDTLTPYFFPAFS 262  
DB 265 LNNABEKKVLSOKIAELSNEIKEAQTQIELVSESGQLKESHVKDRDL-----FS 316  
QY 263 -----KEPMTK-ADLEQCWDIPNFFQLFCNFSVIVESYSETITKMLKAIEDLPKQ 312  
DB 317 LRDIETHQRESSTRVSELEA-----QL-----ESSEQRISDLTYDLKDAEE 359  
QY 313 DKAPDHGGLISKMLPGQDRLGCELQNLSCRFKFKHEKCKQCAHLSEDCPDVPAHLTEL 372  
DB 360 NKA-----ISSKNLEIMDK-----LEQAQNTIKELMDELGEKDKRHEK 398  
QY 373 D-EAIRLVNSVNOQYQILQMTKRKHEDETAIVLEKMRGQFGWVSELANOAPET----- 424  
DB 399 ESELSLVKSADQVADMKQSL-----DNAEEKKMLSO--RILDISNEIQBAQKTIQEH 451  
QY 425 -----ELIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKLPLEES 470  
DB 452 MEESQLKESHGVKRELTGLRDIHEITHQRESSTRLESELTQLKLEQRVVDLSASNA 511  
QY 471 AESSNFIGYVAKALQHFKE 490  
DB 512 EEEKSLSSMILEITDELKQ 531

Search completed: July 2, 2002, 11:49:54  
Job time: 388 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run On: July 2, 2002, 11:44:42 ; Search time 56.51 seconds  
(without alignments)  
972.952 Million cell updates/sec

Title: US-09-722-544A-2MOD

Perfect score: 2632

Sequence: 1 MKIKAEKNEGPRSMWQLHW.....FIGYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :			
A.Geneseq_032802.*			
1:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*	432	16.4
2:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*	405.5	15.4
3:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*	400.5	15.2
4:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*	320	12.2
5:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*	173.5	6.6
6:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*	160.5	6.1
7:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*	159.5	6.1
8:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*	159.5	6.1
9:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*	159.5	6.1
10:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*	159.5	6.1
11:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*	159.5	6.1
12:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*	159.5	6.1
13:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*	159.5	6.1
14:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*	159.5	6.1
15:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*	159.5	6.1
16:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*	159.5	6.1
17:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*	159.5	6.1
18:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*	159.5	6.1
19:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*	159.5	6.1
20:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*	159.5	6.1
21:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*	159.5	6.1
22:	/SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*	159.5	6.1

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES			
Result No.	Score	Query Match Length	ID Description
1	2582	98.1	495 20 AAY30785 Protein encoded by
2	2460	93.5	456 20 AAY30793 Immature human HKN
3	2443	92.8	477 20 AAY30786 Protein encoded by
4	2346	89.1	446 20 AAY30792 Mature secreted hum
5	1753.5	66.6	465 20 AAY30791 Bovine HKNG1 ortho
6	1635	62.1	466 20 AAY30787 Protein encoded by
7	1554	59.0	450 20 AAY30788 Guinea pig HKNG1 o
8	1277	48.5	374 20 AAY30789 Guinea pig HKNG1 o
9	1272.5	48.3	373 20 AAY30790 Guinea pig HKNG1 o
10	925.5	35.2	521 20 AAY30794 Amino acid sequenc
11	434	16.5	449 22 AAU28048 Novel human secret

12	432	16.4	448	12	AA11704	Cytolysis Inhibito
13	405.5	15.4	416	22	ABB50285	Apolipoprotein J o
14	400.5	15.2	446	21	AAE03441	Porcine clusterin
15	320	12.2	363	22	AAE03764	Human gene 1 encod
16	173.5	6.6	148	22	AAO00607	Human polypeptide
17	160.5	6.1	116	21	AAO03745	Human secreted pro
18	159.5	6.1	139	22	AAE03783	Human gene 1 encod
19	159.5	6.1	139	22	AAE01703	Human gene 4 encod
20	152.5	5.8	247	22	AAE03787	Human gene 1 encod
21	149	5.7	1372	19	AAV56473	Protein with Rho p
22	145	5.5	944	21	AAV67600	Human adipose tiss
23	145	5.5	1388	19	AAW56475	Protein with Rho p
24	143	5.4	1931	22	ABW61012	Drosophila melanog
25	141.5	5.4	3433	18	AAW22017	Utrophin. Homo sa
26	140	5.3	934	22	AAU01768	Human secreted pro
27	140	5.3	2517	21	AAV71159	Human phosphodiect
28	139.5	5.3	2688	22	AAW40883	Human polypeptide
29	139	5.3	2633	22	ABG06505	Novel human diagno
30	138.5	5.3	2663	22	AAW39097	Human polypeptide
31	138	5.2	5373	22	AAU14803	Novel bone marrow
32	138	5.2	5447	22	AAU14897	Novel bone marrow
33	135.5	5.1	1427	12	AAE10534	Human 160kD mediat
34	135	5.1	1374	22	AAW69070	Human male enhance
35	133	5.1	990	22	AAW78520	Human protein SEQ
36	132.5	5.0	1411	17	AAW02258	Nucleolar/endosoma
37	131	5.0	816	16	AAW66931	AMML chromosome in
38	130.5	5.0	2482	16	AAW72826	Human mitotin. Ho
39	130.5	5.0	2482	19	AAW23996	Human mitotin amin
40	130	4.9	1788	22	AAW40467	Human polypeptide
41	130	4.9	3248	17	AAW99795	Kinetochore protei
42	129	4.9	885	16	AAW66930	AMML chromosome in
43	129	4.9	931	22	AAW79504	Human protein SEQ
44	128.5	4.9	1392	20	AAW06999	Restin protein seq
45	127.5	4.8	721	21	AAE21227	Protein encoded by

ALIGNMENTS

RESULT 1	
AAV30785	AAV30785 standard; Protein; 495 AA.
ID	AAV30785 standard; Protein; 495 AA.
AC	AAV30785;
XX	
DT	23-NOV-1999 (first entry)
XX	
DE	Protein encoded by human HKNG1 cDNA.
XX	
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;
KW	schizophrenia; splice variant.
XX	
OS	Homo sapiens.
XX	
PN	WO9947535-A1.
XX	
PD	23-SEP-1999.
XX	
PF	16-MAR-1999; 99WO-US05606.
XX	
PR	16-MAR-1998; 98US-0078044.
PR	05-JUN-1998; 98US-0088312.
PR	28-OCT-1998; 98US-0106056.
PR	22-JAN-1999; 99US-0236134.
PA	(MILL-) MILLENNIUM PHARM INC.
PA	(REGC) UNIV CALIFORNIA.
XX	
PI	Chen H, Freilmer NB;
XX	
DR	WPI: 1999-562047/47.
DR	N-PSDB; AA210750.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
PS Claim 1; Fig 1A-B; 205pp; English.  
XX  
CC The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1  
CC is a gene associated with bipolar affective disorder (BAD). HKNG1  
CC polynucleotides are useful to identify compounds modulating HKNG1 gene  
CC expression or HKNG1 polypeptide expression/activity. Compounds inhibiting  
CC or enhancing HKNG1 gene expression or activity in individuals can then  
CC be administered therapeutically to treat HKNG1-mediated disorders,  
CC especially neuropsychiatric disorders e.g. BAD, schizophrenia, or  
CC HKNG1-mediated myopia disorders, such as early-onset autosomal  
CC dominant myopia. The polynucleotides can be used in gene therapy  
CC techniques to treat such disorders. They are also useful in diagnosis  
CC to identify individuals having, or at risk of developing, HKNG1-mediated  
CC disorders due to mutations in the HKNG1 gene. Such mutations especially  
CC result in the production of a protein with a different sequence to  
CC the human full-length HKNG1 polypeptide or splice variant sequences,  
CC especially the substitution of a lysine for a glutamic acid at residue  
CC 202 or 184. The polynucleotides are also useful in gene mapping, to  
CC produce probes or primers to identify similar sequences (e.g. mutants  
CC or sequences from different species) and to produce transgenic  
CC animals.  
XX  
XX Sequence 495 AA;

Query Match 98.1%; Score 2582; DB 20; Length 495;  
Best Local Similarity 99.0%; Pred. No. 1.2e-206;  
Matches 489; Conservative 1; Mismatches 4; Indels 0; Gaps 0;

QY 1 MKTKAEKNGPSRWQLHWDIANNNGNKKPPLLVFIVCLLWLDKSHCAPTWKDKTATS 60  
DB 1 mkkaeknegpsrwqlhwgdianngnkkppllvfivclllwldkshcaptivekdkats 60  
QY 61 ENLKSFSEGEIDAEVVKALTIKIMIMMERKEHTNLMSTLUKCKREEKQEAALKLL 120  
DB 61 enlkfsfsegeidadeevvkaltikimkimmerkehtnlmstllkckreekqealkll 120  
QY 121 NEVOEHLEERLCRESLADSGECRSCLENNCMRIYTTTCOPSSVSKNKKIERFRKIYQ 180  
DB 121 nevgehleeerlcrsladshgecrscleenncmrlyttcqpssvsknkierffrklyq 180  
QY 181 FLPPFHEDNEKDLPISEKLTIEKDAQLTQMEDVFSQLTVDVNSLNFNSFNVRQMQEFDQ 240  
DB 181 flpfhednekdldpisekleedakltqkdedvfsqldvnslnfnsfnvrqkqepdq 240  
QY 241 TFOSHFISDLDLTPYFFPAFSKEPTKADLEOCWDIPNPFQJFCNFSVIYESVETIT 300  
DB 241 tfqshfisdldltpyffpafskcptkadeqwdipnffqjfcnfsvsiyesvsetit 300  
QY 301 KMLKATEDLPKQKADPHGGLISKMLPGDGRGLCGELDONLSRCFKFHEKCKQCAHLSE 360  
DB 301 kmlkatedlpkqkdpghggliskmlpgdgrglcgeldonlsrckfkfhekckqcahlse 360  
QY 361 DCPDVPALHTELDEAIRLVNVSNOQYGOILQMTKRKLEDTAYLVERKMRGQFGWVSELANQ 420  
DB 361 dcpdvpalhteldeairlvnvsnoqyggqilqmtkrkhledtaylvekmrgfgwvselanq 420  
QY 421 APETEIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTKIPLESAESSNFIYV 480  
DB 421 apetEIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTKIPLESAESSNFIYV 480  
QY 481 VAKALQHFKEHFT 494  
DB 481 vakalqhfkehft 494

RESULT 2  
AAY30793

ID AAY30793 standard; Protein; 466 AA.  
XX  
AC AAY30793;  
XX  
DT 23-NOV-1999 (first entry)  
XX  
DE Immature human HKNG1 protein form 2.  
XX  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
PF 16-MAR-1999; 99WO-US05606.  
XX  
PR 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
PI Chen H, Freimer NB;  
PI  
XX  
DR WPI; 1999-562047/47.  
XX  
PT New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
PS Claim 22; Fig 17; 205pp; English.  
XX

The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1 is a gene associated with bipolar affective disorder (BAD). HKNG1 polynucleotides are useful to identify compounds modulating HKNG1 gene expression or HKNG1 polypeptide expression/activity. Compounds inhibiting or enhancing HKNG1 gene expression or activity in individuals can then be administered therapeutically to treat HKNG1-mediated disorders, especially neuropsychiatric disorders e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as early-onset autosomal dominant myopia. The polynucleotides can be used in gene therapy techniques to treat such disorders. They are also useful in diagnosis to identify individuals having, or at risk of developing, HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such mutations especially result in the production of a protein with a different sequence to the human full-length HKNG1 polypeptide or splice variant sequences, especially the substitution of a lysine for a glutamic acid at residue 202 or 184. The polynucleotides are also useful in gene mapping, to produce probes or primers to identify similar sequences (e.g. mutants or sequences from different species) and to produce transgenic animals.

Sequence 466 AA;

Query Match 93.5%; Score 2460; DB 20; Length 466;  
Best Local Similarity 99.8%; Pred. No. 1.7e-196;  
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 30 MKPPLLVFIVCLLWLDKSHCAPTWKDKTALISENLKSFSEVGEIDAEVVKALTGKQMK 89  
DB 1 mkpplllvfivclllwldkshcaptivekdkaltaiselnklsfsevgeldadeevkaltgkqmk 60  
QY 90 IMMERKEHTNLMSTLUKCKREEKQEAALKLLNEVQEHLEERLCRESLADSGECRSL 149  
DB 61 immerkehtnlmstllkckreekqalkllnevqehleerlcrsladsgwgcrcsl 120

Qy	150	ENNCMRIYTTQCPSWSVSKNKNIERFRKRYQOFLPPHEONEKDXLPSEKLIIEKDAQLTQM	209
Db	121	enncmrltyttcqpswsvsknkierfrkyqyflfphednekdipsekleeadaqltqm	180
Qy	210	EDVFSQLTVDVNSLFINRSENVFRMQOEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA	269
Db	181	edvfslqtdvnslnfrsnfnvfrmqeqdqtqshfisdtdltepyffpafskepmtka	240
Qy	270	DLBQCWDIPNFFQOLFNCFSVSYIYESVSEITKMLKAIEDLPQDKAPDHHGLLISKMLPQ	329
Db	241	dleqcwdipnffqlfncfsvisyevsetikmlkaiedlpqdkapdhgglliskmlpq	300
Qy	330	DRGLCGELDONLSRCFKFKCKOKCAHLSDECDDVPALHTELDEAIRLVNVSNOQYGOI	389
Db	301	drglcgeldnlsrckfkfkckckqahlsedcddvpalhhteldeairlvnvsnnqyqgi	360
Qy	390	LQWTRKHLEDTAYLVEKMRQCFGWSELANQAPETELIENSIOQVPRIHEGNISKODETM	449
Db	361	lqmrkhltdatylvekmrgcfgwselanqapetelinsiqvprihegniskodecm	420
Qy	450	MTDLSILPSSNFTLKIPLESASSSNFPIGYVAKALQHFKEHFTW	495
Db	421	mtdslilpssnftlkipleaessnfiqyvvakalqhfkehtw	466

RESULT	3	
AAAY30786		
ID	AAAY30786 standard; Protein; 477 AA.	
XX		
XX	AAAY30786;	
XX		
XX	23-NOV-1999 (first entry)	
DE	Protein encoded by a human HKNG1 splice variant HKNG1-V1.	
XX		
XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;	
KW	schizophrenia; splice variant.	
XX		
XX	Homo sapiens.	
XX		
XX	WO9947535-A1.	
XX		
XX	23-SEP-1999.	
PD		
XX		
XX	16-MAR-1999; 99WO-US05606.	
PF		
XX		
XX	16-MAR-1998; 98US-0078044.	
PR		
PR	05-JUN-1998; 98US-0088312.	
PR	28-OCT-1998; 98US-0106056.	
PR	22-JAN-1999; 99US-0236134.	
XX		
PA	(MILL-) MILLENNIUM PHARM INC.	
PA	(REGC) UNIV CALIFORNIA.	
XX		
XX	Chen H, Freimer NB;	
PI		
XX		
XX	WPI; 1999-562047/47.	
DR		
DR	N-PSDB; AAZ10751.	
XX		
XX		
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of	
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and	
PT	schizophrenia	
XX		
XX	Claim 1; Fig 2A-B; 205pp; English.	
FS		
XX		
CC	The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)	
CC	splice variant. HKNG1 is a gene associated with bipolar affective	
CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compound	
CC	modulating HKNG1 gene expression or enhancing HKNG1 polypeptide expression/	
CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression or	
CC	activity in individuals can then be administered therapeutically to	
CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders	
CC		

e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as early-onset autosomal dominant myopia. The polynucleotides can be used in gene therapy techniques to treat such disorders. They are also useful in diagnosis to identify individuals having, or at risk of developing, HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such mutations especially result in the production of a protein with a different sequence to the human full-length HKNG1 polypeptide or splice variant sequences, especially the substitution of a lysine for a glutamic acid at residue 202 or 184. The polynucleotides are also useful in gene mapping, to produce probes or primers to identify similar sequences (e.g. mutants or sequences from different species) and to produce transgenic animals.

Sequence 477 AA;

Query Match	92.88;	Score	2443;	DB	20;	Length	477;		
Best Local Similarity	98.96;	Pred. No.	4.5e-195;						
Matches	465;	Conservative	2;	Mismatches	3;	Indels	0;	Gaps	0;
Qy	25	NNSGNMKPLLVFVIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGETADDEAEVKKALTG	84						
Db	7	snsngnmkpllvfivclllwkdshcaptwkdktaisenlksfsevgeidaeevkkaltg	66						
Qy	85	IKQMKIMMERKEKHTNLMSTLTKCREBKQBAKLKLLNEVQEHLEPEERLCRESLADSWGE	144						
Db	67	lkqmkimmerkekehtnlmstllkcrekqgalkllnevqehleeeerlcresladswe	126						
Qy	145	CRSCLENNCMRLTYTCOPSWSSVKNKIERFRPKYQFLFPFHENEKDKLPISSEKLEKDA	204						
Db	127	crsclenncmrltytcqpswsvknkierfrkiyqflfpfhednekdlpisekleeda	186						
Qy	205	QLTQMEDVFSQLTVDVNSLFNRFSNVFQMOQEFDTQSFHISDTDJTEYFFPFAFSKE	264						
Db	187	qltqmedvfgltvdvnslnfrsnfvrmqgdfqdtfqskfisdtdltepyffpafske	246						
Qy	265	PMTKADLEQCDNDIPNFOLFNFVSIVSYESVETITKMLKAIEDLPKODKAPDHGGGLSK	324						
Db	247	pmtkadlegcdwipwffglfcnfsvsvsyvsetitkmlkaiedlpqgdkapdhggllsk	306						
Qy	325	MLPGDRLGCGLEDQNLSCRPFHKCKQCAHLSEDCPDVPALHTELDEAIRLVNSNQ	384						
Db	307	mlpgdrglcgeldqnlscrcfkmekckqahisedcpdvpalhteleairlvnsnq	366						
Qy	385	OXYGTOILQWTRKHLEDYALVEKMRGQFGWSELANQAPETETIFNISOVVPRIHEGNISK	444						
Db	367	qyvgllqtmrkhnledyalivekmrgqfgvwselanqapetelifnslqvvprihegnisk	426						
Qy	445	QDETMMTDLISLPSSNFTLTKIPLESAESSNFIPIGVVAKALQHFKEHKPT	494						
Db	427	qdetmmtdlislpssnftlkipleesaessnfigvvvakalchfkehkt	476						

RESULT	4	
AAAY30792		
ID	AAAY30792 standard; Protein; 446 AA.	
XX		
AC	AAAY30792;	
XX		
DT	23-NOV-1999 (first entry)	
XX		
DE	Mature secreted human HKNG1 protein sequence.	
XX		
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;	
KW	schizophrenia; splice variant.	
XX		
OS	Homo sapiens.	
XX		
PN	W09947535-A1.	
XX		
PD	23-SEP-1999.	
XX		

PF	16-MAR-1999;	99WO-US05606.	
XX	16-MAR-1998;	98US-0078044.	
PR	05-JUN-1998;	98US-0088312.	
PR	28-OCT-1998;	98US-0106056.	
PR	22-JAN-1999;	99US-0236134.	
XX	(MILL-) MILLENNIUM PHARM INC.		
PA	(REGC ) UNIV CALIFORNIA.		
XX	Chen H, Freimer NB;		
XX	WPI; 1999-562047/47.		
XX	New HKNG1 polynucleotides useful in diagnosis and treatment of		
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and		
PT	schizophrenia		
XX	Claim 5; Fig 17; 205pp; English.		
XX	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1		
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1		
CC	polynucleotides are useful to identify compounds modulating HKNG1 gene		
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting		
CC	or enhancing HKNG1 gene expression or activity in individuals can then		
CC	be administered therapeutically to treat HKNG1-mediated disorders,		
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or		
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal		
CC	dominant myopia. The polynucleotides can be used in gene therapy		
CC	techniques to treat such disorders. They are also useful in diagnosis		
CC	to identify individuals having, or at risk of developing, HKNG1-mediated		
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially		
CC	result in the production of a protein with a different sequence to		
CC	the human full-length HKNG1 polypeptide or splice variant sequence,		
CC	especially the substitution of a lysine for a glutamic acid at residue		
CC	202 or 184. The polynucleotides are also useful in gene mapping, to		
CC	produce probes or primers to identify similar sequences (e.g. mutants		
CC	or sequences from different species) and to produce transgenic		
CC	animals.		
XX	Sequence 446 AA;		
SQ			
Query Match 89.1%; Score 2346; DB 20; Length 446;			
Best Local Similarity 99.8%; Pred. No. 5e-187;			
Matches 445; Conservative 1; Mismatches 0; Indels 0; Gaps 0;			
QY	50	APTWKDTAISENLKSFSEVGIEDADEEYKALTGTGKIMMERKEKEHTLMSTLKKC	109
DB	1	aptwkdtaislenkfsfsevgiedadeevkaltgikqmkimmerkekehtlmstlkkc	60
QY	110	REEKQALKLLNEVQHLEERLCHRESLADSWGECRSCLENNCMRIYTCQPSWSSVKN	169
DB	61	reekqalkllnevqhleeerlchresladswgecrsclenncmriyttcqpssvskn	120
QY	170	KIERFRKIYQFLFPHEHNEKDLPTSEKLEIKDAQLTQMEDVFSQTLVDVNSLNRSPN	229
DB	121	kierfrkiyqflfphehnekdplsekleiedaqldqmedvfsqtlvdvnslnrskn	180
QY	230	VFRMQQEDTQTFQSHFISDTLTTEYFPFAPFSKEPMTKADLEQCDWIPNFFQLCFNESV	289
DB	181	virmqqeqdqtqsfhsfisdtdlteyffpafskpmtkadleqcdwipnffqlcfntsv	240
QY	290	SVYESVSEITKMLKAIEDLPQDKAPDHGGLISKMLPCQDRGLCGELDONLSRCFKPHE	349
DB	241	svyesvseitkmlkaiedlpqdkapdhggllskmlpqdrglgeidqnlsrckfthe	300
QY	350	KCKQCOAHLSEDCPDVPAHLTDLDEAIRLVNVSNOQYGOILQMTKRKHLEDYAYLVEKMRG	409
DB	301	kcqkcahlsecdpvpahltdeldeairlvnvsnoqyqilqmtkrkhledtaylvekmrg	360
QY	410	QFGWVSELANQAPETEIFNSIQVPRTHEGNISQKDETMWTDLSILPSSNFTLKIPLEE	469

Db	361	qfgwvselanqapetelifnsiqvprthegniskqdetmtdlsilpsnftlkiplee	420
QY	470	SAESSNFIGVYVAKALQHKFKEFTW	495
DB	421	saessnfigyvvakalqhkfhkftw	446
RESULT	5		
AAV30791			
ID	AAV30791	standard; Protein; 465 AA.	
XX	XX	AAV30791;	
AC	XX		
XX	23-NOV-1999	(first entry)	
DT	XX	Bovine HKNG1 ortholog splice variant protein.	
DE	XX		
XX	XX	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
KW	XX	neuropsychiatric disorder; early-onset autosomal dominant myopia;	
KW	XX	schizophrenia; splice variant.	
OS	XX	Bos sp.	
XX	XX	WO9947535-A1.	
PN	XX	23-SEP-1999.	
PD	XX	16-MAR-1999; 99WO-US05606.	
XX	XX	16-MAR-1998; 98US-0078044.	
PR	XX	05-JUN-1998; 98US-0088312.	
PR	XX	28-OCT-1998; 98US-0106056.	
PR	XX	22-JAN-1999; 99US-0236134.	
XX	XX	(MILL-) MILLENNIUM PHARM INC.	
PA	XX	(REGC ) UNIV CALIFORNIA.	
XX	XX	Chen H, Freimer NB;	
PI	XX	WPI; 1999-562047/47.	
DR	XX	N-PSDB; AAZ10759, AAZ10760, AAZ10761.	
DR	XX	New HKNG1 polynucleotides useful in diagnosis and treatment of	
PT	XX	neuropsychiatric disorders, e.g. bipolar affective disorders and	
PT	XX	schizophrenia	
PS	XX	Claim 1; Fig 11A-B; 205pp; English.	
XX	XX	The present sequence is encoded by bovine HKNG1 ortholog splice	
CC	XX	variant. HKNG1 (Hong Kong new gene 1) is a gene associated	
CC	XX	with bipolar affective disorder (BAD). HKNG1 polynucleotides are	
CC	XX	useful to identify compounds modulating HKNG1 gene expression or	
CC	XX	HKNG1 polypeptide expression/activity. Compounds inhibiting or	
CC	XX	enhancing HKNG1 gene expression or activity in individuals can	
CC	XX	then be administered therapeutically to treat HKNG1-mediated	
CC	XX	disorders, especially neuropsychiatric disorders e.g. BAD,	
CC	XX	schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	XX	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	XX	in gene therapy techniques to treat such disorders. They are also useful	
CC	XX	in diagnosis to identify individuals having, or at risk of developing,	
CC	XX	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	XX	mutations especially result in the production of a protein with a	
CC	XX	different sequence to the human full-length HKNG1 polypeptide or	
CC	XX	splice variant sequences, especially the substitution of a lysine for	
CC	XX	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	XX	useful in gene mapping, to produce probes or primers to identify	
CC	XX	similar sequences (e.g. mutants or sequences from different species)	
CC	XX	and to produce transgenic animals.	
XX	XX	Sequence 465 AA;	
SQ			

Query Match 66.6%; Score 1753.5; DB 20; Length 465;

Best Local Similarity 70.8%; Pred. No. 1.3e-137;  
Matches 330; Conservative 58; Mismatches 77; Indels 1; Gaps 1;

QY 30 MKPPLLVIIVCLLWKDQSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGIKQMK 89  
DQ 1 mkppllviivllqrdccqaptgdrtsredpkgsageldvdeevkalligmqmk 60  
QY 90 IMMERKEKEHTNLMTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSCL 149  
DQ 61 ilmerreehsklmrtikckreekqalklmevqehleeeerlcqyslmgswdeckscl 120  
QY 150 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKEDAQLTQM 209  
DQ 121 esdcmrlyttcqswwsmkstierfrkiyqflpfphedekelpvgekftedvqlmqi 180  
QY 210 EDVFSQLTVDVNSLFRNSFNVRMQQEFDTQFOSHFISDLDLTPYFFPAFSKEPMTKA 269  
DQ 181 envfsqldvvgflymsfhwfkqmqqefdlafqsfmsdtdsmepfyfpafskepaka 240  
QY 270 DLEQCWDIPNPFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 329  
DQ 241 hpmqswdipfifcfnfslyvqsvsatvttemikaledlskqkdsahgqpssttvpvr 300  
QY 330 DRGLCGELDQNLSCRFKHEKQKCOAHLSEDCPDVPAHTELDLTPYFFPAFSKEPMTKA 389  
DQ 301 grgicgepgqnsseclgharcqkcdylwadcpavpelytkadealelvnlsnqqvayq 360  
QY 390 LQMTRKHLEDYALVEKMRGQFGWVSELANOAPETEIFNSIQVVPRTHEGNIKQDETM 449  
DQ 361 lqmtqhledtlylmeqnregfwwtelasqtpgsenlfsfkwvpgvghnegnsqddq 420  
QY 450 MTDLSILPSSNFTLIKIPLESSESNFTGYVAKALQHEKHFXTW 495  
DQ 421 i-disilpssnftitiplesseassdfisymlakavghkfhkfw 465

RESULT 6  
AAY30787  
ID AAY30787 standard; Protein; 466 AA.  
XX AAY30787;  
XX AAY30787;  
XX AAY30787;  
XX 23-NOV-1999 (first entry)  
XX Protein emoded by the guinea pig HKNG1 ortholog gphkng1815.  
XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
XX neuropsychiatric disorder; early-onset autosomal dominant myopia;  
XX schizophrenia; splice variant.  
XX Cavia cobaya.  
OS  
XX WO9947535-A1.  
XX  
XX 23-SEP-1999.  
XX  
XX 16-MAR-1999; 99WO-US05606.  
XX  
XX 16-MAR-1998; 98US-0078044.  
XX  
XX 05-JUN-1998; 98US-0088312.  
XX  
XX 28-OCT-1998; 98US-0106056.  
XX  
XX 22-JAN-1999; 99US-0236134.  
XX  
XX (MILL-) MILLENNIUM PHARM INC.  
XX (REGC ) UNIV CALIFORNIA.  
XX  
XX Chen H, Freimer NB;  
XX  
XX WPT; 1999-562047/47.  
XX  
XX N-PSDB; AA210755.  
XX  
XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
XX neuropsychiatric disorders, e.g. bipolar affective disorders and

PT schizophrenia -  
XX  
PS Claim 1; Fig 7A-B; 205pp; English.  
XX  
CC The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.  
CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective  
CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
CC activity in individuals can then be administered therapeutically to  
CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia. The polynucleotides can be used  
CC in gene therapy techniques to treat such disorders. They are also useful  
CC in diagnosis to identify individuals having, or at risk of developing,  
CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
CC mutations especially result in the production of a protein with a  
CC different sequence to the human full-length HKNG1 polypeptide or  
CC splice variant sequences, especially the substitution of a lysine for  
CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
CC useful in gene mapping, to produce probes or primers to identify  
CC similar sequences (e.g. mutants or sequences from different species)  
CC and to produce transgenic animals.  
XX  
SQ Sequence 466 AA;

Query Match 62.1%; Score 1635; DB 20; Length 466;  
Best Local Similarity 66.5%; Pred. No. 9.5e-128;  
Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;

QY 30 MKPPLLVIIVCLLWKDQSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALTGIKQMK 89  
DQ 1 mkppllviivclllwkdcaptwkdktaisenansfseageidvgevkialgikgm 60  
QY 90 IMMERKEKEHTNLMTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSCL 149  
DQ 61 immerreehsklmrtikckreekqalklmevqehleeeeslcqysladsweclacl 120  
QY 150 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIKEDAQLTQM 209  
DQ 121 esncmrlyttcqswwsmkstierfrkiyqflpfphedekelpvgekftedvqlmqi 179  
QY 210 EDVFSQLTVDVNSLFRNSFNVRMQQEFDTQFOSHFISDLDLTPYFFPAFSKEPMTKA 269  
DQ 180 ehvfsqldsvtsifnrslyvfkqlrrefdaqfsgtdvtepfifpslsepaya 239  
QY 270 DLEQCWDIPNPFQFCNFSVSIYESVETITKMLKAIEDLPKQKAPDHGGLISKMLPGQ 329  
DQ 240 daepswaiipnvfqlcnlsfsvyqsvseklittratedppkqkdsngggsakilpeq 299  
QY 330 DRGLCGELDQNLSCRFKHEKQKCOAHLSEDCPDVPAHTELDLTPYFFPAFSKEPMTKA 389  
DQ 300 drgsdglgqnlscdvnfkrccqkcdyisddcpnvpeylrelnealfrsrsnqqyqv 359  
QY 390 LQMTRKHLEDYALVEKMRGQFGWVSELANOAPETEIFNSIQVVPRT--HEGNIKQDE 447  
DQ 360 vqmtqyhledtllmekmregfwwselayqsgaedifnpkvkmvalshnegnsdqd 419  
QY 448 TMTDLSILPSSNFTLIKIPLESSESNFTGYVAKALQHEKHFXTW 495  
DQ 420 tvvps-silpssnftlsspleksagnanfidhvvvekvqlghkfhkfw 466

RESULT 7  
AAY30788  
ID AAY30788 standard; Protein; 450 AA.  
XX AAY30788;  
XX AAY30788;  
XX 23-NOV-1999 (first entry)  
XX Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.





CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such mutations especially result in the production of a protein with a different sequence to the human full-length HKNG1 polypeptide or splice variant sequences, especially the substitution of a lysine for a glutamic acid at residue 202 or 184. The polynucleotides are also useful in gene mapping, to produce probes or primers to identify similar sequences (e.g. mutants or sequences from different species)

XX CC and to produce transgenic animals.

SQ Sequence 374 AA;

Query Match 48.3%; Score 1277; DB 20; Length 374;  
Best Local Similarity 54.5%; Pred. No. 4.8e-98;  
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;

Qy 30 MKPPLLVIIVCLLWKDHCAPTWKDKTATSENLSKSFSEVGEIDADDEVKALGTGKQMK 89  
Db 1 mkpllpfvcllwkdchcptwkdktaisenansiseageldvdevkialigkqmk 60

Qy 90 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOEHLERLCSLADSWGECRSCL 149  
Db 61 immerreehsklmktlkkckeeqalklmnevheleeeslcqvsldswdecrac 120

Qy 150 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLPFFHEDNEKDLPISEKLIKDAQLTQM 209  
Db 121 esncmrfdttcqpawssvkn----- 140

Qy 210 EDVFSQLTVDVNSLNFNSFNVRQMQOEFDTQFSHFISDLDLTPYFFPAFSKEPTKA 269  
Db 141 -----mepayra 147

Qy 270 DLEOCWDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329  
Db 148 daepswaipnvqllcnlsfsvysvseklittratedppkqkdsngqgppiskilpeq 207

Qy 330 DRGLCGELDQNLSCRFKHEKQKQAHLSDECDPVPALHTLDEAIRLVNVSQOYQGI 389  
Db 208 drgsdgklgnlscvnrfrckqkqdylsddcpnvpeylrelnealrlvsrsnqydvq 267

Qy 390 LQMTKRKHEDTAYLVEKMRGQGWVSELANOAPETEIFNSIOVVPRI--HEGNISKODE 447  
Db 268 vqmtcqhlydettllmekmreqfgwvselayspgaedfnpvmvalsahegnssoddd 327

Qy 448 TMTDLSLPSNFTLKIPLSESSNFYGYVYVAKALQHFKEFTKW 495  
Db 328 tvvps-sllpssntflspleksagnanfidhvvekvqlghfkehftkw 374

RESULT 9  
ID AAY30790  
XX AC AAY30790 standard; Protein; 373 AA.  
XX AC AAY30790;  
XX DT 23-NOV-1999 (first entry)  
XX DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d.  
XX KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
XX KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
XX KW schizophrenia; splice variant.  
XX OS Cavia cobaya.  
XX PN W09947535-A1.  
XX PD 23-SEP-1999.  
XX PF 16-MAR-1999; 99WO-US05606.  
XX PF 16-MAR-1998; 98US-0078044.  
XX PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX (MILL-) MILLENNIUM PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX Chen H, Freimer NB;  
PI WPI; 1999-562047/47.  
XX DR N-PSDB; AAZ10758.  
XX New HKNG1 polynucleotides useful in diagnosis and treatment of neuropsychiatric disorders, e.g. bipolar affective disorders and schizophrenia -  
PS Claim 1; Fig 10A-B; 205pp; English.  
XX The present sequence is encoded by a guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective disorder (BAD). HKNG1 polynucleotides are useful to identify compounds modulating HKNG1 gene expression or HKNG1 polypeptide expression/activity. Compounds inhibiting or enhancing HKNG1 gene expression or activity in individuals can then be administered therapeutically to treat HKNG1-mediated disorders, especially neuropsychiatric disorders e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as early-onset autosomal dominant myopia. The polynucleotides can be used in gene therapy techniques to treat such disorders. They are also useful in diagnosis to identify individuals having, or at risk of developing, HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such mutations especially result in the production of a protein with a different sequence to the human full-length HKNG1 polypeptide or splice variant sequences, especially the substitution of a lysine for a glutamic acid at residue 202 or 184. The polynucleotides are also useful in gene mapping, to produce probes or primers to identify similar sequences (e.g. mutants or sequences from different species) and to produce transgenic animals.

SQ Sequence 373 AA;

Query Match 48.3%; Score 1272.5; DB 20; Length 373;  
Best Local Similarity 54.5%; Pred. No. 1.1e-97;  
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

Qy 30 MKPPLLVIIVCLLWKDHCAPTWKDKTATSENLSKSFSEVGEIDADDEVKALGTGKQMK 89  
Db 1 mkpllpfvcllwkdchcptwkdktaisenansiseageldvdevkialigkqmk 60

Qy 90 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOEHLERLCSLADSWGECRSCL 149  
Db 61 immerreehsklmktlkkckeeqalklmnevheleeeslcqvsldswdecrac 120

Qy 150 ENNCMRIYTTCCPSWSSVKNKIERFRKIYQFLPFFHEDNEKDLPISEKLIKDAQLTQM 209  
Db 121 esncmrfdttcqpawssvkn----- 140

Qy 210 EDVFSQLTVDVNSLNFNSFNVRQMQOEFDTQFSHFISDLDLTPYFFPAFSKEPTKA 269  
Db 141 -----mepayra 146

Qy 270 DLEOCWDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQKAPDHPGGLISKMLPGQ 329  
Db 147 daepswaipnvqllcnlsfsvysvseklittratedppkqkdsngqgppiskilpeq 206

Qy 330 DRGLCGELDQNLSCRFKHEKQKQAHLSDECDPVPALHTLDEAIRLVNVSQOYQGI 389  
Db 207 drgsdgklgnlscvnrfrckqkqdylsddcpnvpeylrelnealrlvsrsnqydvq 266

Qy 390 LQMTKRKHEDTAYLVEKMRGQGWVSELANOAPETEIFNSIOVVPRI--HEGNISKODE 447  
Db 267 vqmtcqhlydettllmekmreqfgwvselayspgaedfnpvmvalsahegnssoddd 326

Qy	448	TWMTDLSTLPSNTLTKLPLEBESSESNFYGVVAKALQHFKFTW	495
Db	327	tvpbs-silpsnftlspksagnanfidhvkvqlghfkfktw	373
RESULT 10			
AAU30794	ID	AAU30794 standard; Protein; 521 AA.	
XX	AC	AAU30794;	
XX	AC		
XX	DT	23-NOV-1999 (first entry)	
XX	DE	Amino acid sequence of human HKNG1 splice variant	HKNG1-delta7.
XX	KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;	
XX	KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;	
XX	KW	schizophrenia; splice variant.	
OS	XX	Homo sapiens.	
PN	XX	WO9947535-A1.	
XX	PD	23-SEP-1999.	
XX	PF	16-MAR-1999; 99WO-US05606.	
XX	PR	16-MAR-1998; 98US-0078044.	
XX	PR	05-JUN-1998; 98US-0088312.	
PR	XX	28-OCT-1998; 98US-0106056.	
PR	XX	22-JAN-1999; 99US-0236134.	
PA	PA	(MILL-) MILLENNIUM PHARM INC.	
PA	PA	(REGC ) UNIV CALIFORNIA.	
PI	XX	Chen H, Freimer NB;	
XX	DR	WPI; 1998-562047/47.	
XX	N-	N-PSDB; AAZ10762.	
XX	PT	New HKNG1 polynucleotides useful in diagnosis and treatment of	
XX	PT	neuropsychiatric disorders, e.g. bipolar affective disorders and	
XX	PT	schizophrenia -	
XX	XX	Claim 1; Fig 18A-B; 205pp; English.	
XX	XX	The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)	
CC	CC	splice variant. HKNG1 is a gene associated with bipolar affective	
CC	CC	disorder (BAD). HKNG1 polynucleotides are useful to identify compounds	
CC	CC	modulating HKNG1 gene expression or HKNG1 polypeptide expression/	
CC	CC	activity. Compounds inhibiting or enhancing HKNG1 gene expression/	
CC	CC	activity in individuals can then be administered therapeutically to	
CC	CC	treat HKNG1-mediated disorders, especially neuropsychiatric disorders	
CC	CC	e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as	
CC	CC	early-onset autosomal dominant myopia. The polynucleotides can be used	
CC	CC	in gene therapy techniques to treat such disorders. They are also useful	
CC	CC	in diagnosis to identify individuals having, or at risk of developing,	
CC	CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such	
CC	CC	mutations especially result in the production of a protein with a	
CC	CC	different sequence to the human full-length HKNG1 polypeptide or	
CC	CC	splice variant sequences, especially the substitution of a lysine for	
CC	CC	a glutamic acid at residue 202 or 184. The polynucleotides are also	
CC	CC	useful in gene mapping, to produce probes or primers to identify	
CC	CC	similar sequences (e.g. mutants or sequences from different species)	
CC	CC	and to produce transgenic animals.	
XX	XX	Sequence 521 AA;	
Query Match		35.2%; Score 925.5; DB 20;	Length 521;
Best Local Similarity		77.0%; Pred. No. 1.5e-68;	
Matches 181; Conservative		7; Mismatches 16;	Indels 31; Gaps 3;

CC creating transgenic animals useful for studying the in vivo activities of  
CC the polypeptide as well as for studying modulators of the polypeptides.  
CC (1) induces the proliferation of neural cells and regeneration of nerve  
CC and brain tissue and is useful for the treatment of central and  
CC peripheral nervous system diseases and neuropathies, such as Alzheimer's,  
CC Parkinson's disease, Huntington's disease, and amyotrophic lateral  
CC sclerosis. In addition, (1) is involved in chemotactic or chemokinetic  
CC activity, regulation of haematopoiesis and is useful for treating myeloid  
CC or lymphoid cell disorders, platelet disorders such as thrombocytopenia  
CC and for regeneration of bone, cartilage, tendon, ligament and/or nerve  
CC tissue growth, and in tissue repair, healing of burns, incisions,  
CC ulcers, for treating osteoporosis, osteoarthritis, bone degenerative  
CC disorders, or periodontal disease. Furthermore, (1) is also useful for  
CC gut protection or regeneration and treatment of lung or liver fibrosis,  
CC reperfusion injury in various tissues, various immune deficiencies and  
CC disorders including severe combined immunodeficiency (SCID), bacterial or  
CC fungal infections, autoimmune disorders e.g. multiple sclerosis,  
CC rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic  
CC reactions and conditions, such as asthma or other respiratory problems.  
CC In addition, (1) affects biorhythms or circadian cycles of rhythms,  
CC fertility, metabolism, catabolism, anabolism, storage or elimination of  
CC dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides  
CC analgesic effects or other pain reducing effects, immunoglobulin like  
CC activity and can act as an antigen in a vaccine composition to raise an  
CC immune response. AAU28020-AAU28395 represent novel human secreted protein  
CC amino acid sequences of the invention.

XX Sequence 448 AA;

Query Match 16.5%; Score 434; DB 22; Length 449;

Best Local Similarity 25.8%; Pred. No. 1.1e-27;

Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLLYFIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALT 83

Db 1 mmtkllilfvgl-----twesqvgldgtvsnelqemsnqgskynkeiqnavn 51

QY 84 GIKQMKMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHEERLCRESLADSWG 143

Db 52 gvkikltneerktllsnleaaakkkedalnetresetkklpgvcnetmmalwe 111

QY 144 ECRSCLENNCRIVT-TCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIK 202

Db 112 eckpclkqcmkfyarvcrgsglvgvrgleeflnqspgyfwmgdr-----idsllen 165

QY 203 DAQLTQMEDV----FSQLTVDVNSLNFNRSNVFMQOEFTQTFQSHFIS-DTDLTEPYF 257

Db 166 drqqtldvmqdhfrassildefqdrf-----ftrepqdy--hlypfsphrrphf 218

QY 258 PPAFSKPEMTKADLE-QCWDIPNPFQFCNFSVSIYESVSETITKMLKATEDLPKQKAP 316

Db 219 f--fpksrivrslmpfpyepelnhamqpflemihe-----qqamdlfhspaf 267

QY 317 DHGGLSKMLPGQDGLGELDQNLRCFKFHEKQRCQAHLSDC----PDVPALHT 372

Db 268 qhpptefiregddrtvcvirehnstgclrmkdcqckoreillsvdcstnnpqaklrel 327

QY 373 DEATRLVNSVNOQYQIILQMTKRIKLEDTAYLVKMRQFGVHSELANOAPETEIFNSIQ 432

Db 328 deslqvaerltrkynellskysqkmlntssilleqneqfnwvrslianltdgeddyirvt 387

QY 433 VVPRITHEGNISKQDETMTDLSILPSSNFTLKIPLIESAESSNFIGVVAKALQHF-KEH 491

Db 388 tvas-htsdsdpsvgvtevvvklfidsdpitvtpvevsvrknkpfmetvaekalqeyrkhh 446

RESULT 12

AA11704

ID AAR11704 standard; Protein; 448 AA.

XX

AC AAR11704;

XX

DT 20-JUN-1991 (first entry)  
XX Cytolysis Inhibitor.  
DE Cytolysis Inhibitor.  
XX  
KW cytotoxic inhibitor; perforin; immunological effector molecule;  
infertility.  
XX  
OS Homo sapiens.

XX Key Location/Qualifiers  
FT Peptide 1..21  
FT Protein /label= signal peptide  
FT Protein 22..226  
FT Protein /label= A-chain  
FT Protein 227..448  
FT Protein /label= B-chain

XX DE3933850-A.

XX 18-APR-1991.

XX 06-OCT-1989; 89DE-3933850.

XX 06-OCT-1989; 89DE-3933850.

XX (SCHD ) SCHERING AG.

XX Tschopp J, Jenne D;

XX WPI; 1991-118338/17.

XX DNA sequence coding for cytotoxic inhibitor - is strong inhibitor

XX of terminal complement protein, eg perforin secreted by killer

XX cells

XX Claim 13; Page 9; 15pp; German.

XX This cytotoxic inhibitor is encoded by a 1.7kb BamHI-KpnI fragment

XX isolated from a liver-specific cDNA library. It is a blood plasma

XX component that inhibits immunological effector molecules. It is used

XX for systemic or local treatment of inflammatory or autoimmune

XX diseases mediated by complement or killer cells. It can also be

XX used for detoxification of membrane-active and cytolytic proteins

XX released by bacteria, fungi and insect venoms. The protein has a

XX further use in the treatment of infertility caused by deficiency of

XX cytotoxic inhibitor. Monoclonal antibodies directed against the

XX cytotoxic inhibitor are also covered by the invention. They are used

XX to isolate or quantify the natural protein in human plasma.

XX See also AAQ11501 and AAQ11502.

XX Sequence 448 AA;

XX Query Match 16.4%; Score 432; DB 12; Length 448;

XX Best Local Similarity 25.8%; Pred. No. 1.1e-27;

XX Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 34 LLFVIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTGIQ 87

Db 4 lllfvgl-----twesqvgldgtvsnelqemsnqgskynkeiqnavngvk 54

QY 88 MKIMMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHEERLCRESLADSWGECRS 147

Db 55 iktlietneerktllsnleaaakkkedalnetresetkklpgvcnetmmalweckp 114

QY 148 CLENNCRIVT-TCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLIKDAQ 206

Db 115 clkqcmkfyarvcrgsglvgvrgleeflnqspgyfwmgdr-----idsllendrqq 168

QY 207 TOMEDV----FSQLTVDVNSLNFNRSNVFMQOEFTQTFQSHFIS-DTDLTEPYFP 261

Db 169 thmidvmqdhfrassildefqdrf-----ftrepqdy--hlypfsphrrphff--f 219







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:48:57 ; Search time 23.96 Seconds  
(without alignments)  
504.619 Million cell updates/sec

Title: US-09-722-544A-2MOD  
Perfect score: 2632  
Sequence: 1 MKIAEKNEGSRWQLHW.....FTGVYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued\_Patents\_AA:\*

- 1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*
- 2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*
- 3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*
- 4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*
- 5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*
- 6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2628	99.8	495	4	US-09-268-992-2
2	2460	93.5	466	4	US-09-268-992-64
3	2346	89.1	446	4	US-09-268-992-51
4	2331	88.6	477	4	US-09-268-992-4
5	1751.5	66.5	465	4	US-09-268-992-49
6	1635	62.1	466	4	US-09-268-992-39
7	1554	59.0	450	4	US-09-268-992-41
8	1277	48.5	374	4	US-09-268-992-43
9	1272.5	48.3	373	4	US-09-268-992-45
10	921	35.0	208	4	US-09-268-992-67
11	149	5.7	1388	2	US-08-685-576-1
12	145	5.5	1388	2	US-08-685-576-4
13	131	5.0	816	2	US-08-533-306A-6
14	131	5.0	816	2	US-08-742-923A-6
15	130.5	5.0	2462	1	US-08-328-254-6
16	130	4.9	3248	1	US-08-353-700-1
17	130	4.9	3248	5	PCT-US95-16216-1
18	129	4.9	885	2	US-08-533-306A-4
19	129	4.9	885	2	US-08-742-923A-4
20	126.5	4.8	1354	3	US-08-685-871-2
21	123	4.7	1618	1	US-07-853-913-4
22	122	4.6	1886	4	US-08-938-105-3
23	121	4.6	828	2	US-08-993-228-21
24	119.5	4.5	435	2	US-08-531-439B-4
25	119.5	4.5	1786	4	US-08-973-462-8
26	119.5	4.5	3111	2	US-08-460-309-4
27	119.5	4.5	3111	2	US-08-125-077-4

28	117	4.4	1098	4	US-08-923-992A-8	Sequence 8, Appli
29	117	4.4	1104	4	US-08-923-992A-4	Sequence 4, Appli
30	117	4.4	1939	4	US-09-310-187A-1	Sequence 1, Appli
31	117	4.4	2052	3	US-09-045-201A-2	Sequence 2, Appli
32	116.5	4.4	1057	4	US-09-541-782-10	Sequence 10, Appli
33	116	4.4	1164	4	US-08-923-992A-2	Sequence 2, Appli
34	116	4.4	2154	2	US-08-841-349-4	Sequence 4, Appli
35	115.5	4.4	452	2	US-08-686-599A-18	Sequence 18, Appli
36	115.5	4.4	493	2	US-08-686-599A-5	Sequence 5, Appli
37	115.5	4.4	493	2	US-08-686-599A-16	Sequence 16, Appli
38	114	4.3	1128	4	US-08-923-992A-6	Sequence 6, Appli
39	113.5	4.3	467	2	US-08-686-599A-17	Sequence 17, Appli
40	111.5	4.2	1066	4	US-09-541-782-8	Sequence 8, Appli
41	109	4.1	976	4	US-09-104-324B-4	Sequence 4, Appli
42	109	4.1	1093	5	PCT-US93-03077-1	Sequence 1, Appli
43	108	4.1	2285	4	US-09-308-375-2	Sequence 2, Appli
44	107	4.1	1184	4	US-08-923-992A-10	Sequence 10, Appli
45	107	4.1	1497	1	US-08-623-679-7	Sequence 7, Appli

#### ALIGNMENTS

#### RESULT 1

US-09-268-992-2  
; Sequence 2, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freilmer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; LENGTH: 495  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-2

Query Match 99.8%; Score 2628; DB 4; Length 495;

Best Local Similarity 99.8%; Pred. No. 7.1e-245;

Matches 494; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MKIAEKNEGSRWQLHWGDIANNNGNMPPLLVFTVCLLWLDKSHCATPTWKDKTAIS	60
Db	1	MKIAEKNEGSRWQLHWGDIANNNGNMPPLLVFTVCLLWLDKSHCATPTWKDKTAIS	60
Qy	61	ENLKSFSVEGIDAEVKKALGTGKIMMERKEHTNLMSTLKKCREKEQALKLL	120
Db	61	ENLKSFSVEGIDAEVKKALGTGKIMMERKEHTNLMSTLKKCREKEQALKLL	120
Qy	121	NEVEHLEEEERLCRESLADSWGECRSCLENNCMRIYTCOPSSVKNKIERFRKIYQ	180
Db	121	NEVEHLEEEERLCRESLADSWGECRSCLENNCMRIYTCOPSSVKNKIERFRKIYQ	180
Qy	181	FLFFPHEDNEKDLPTSEKLEDAQLTQMEDVFSQLTVDVNSLNRSNVFRQMOQEFQ	240
Db	181	FLFFPHEDNEKDLPTSEKLEDAQLTQMEDVFSQLTVDVNSLNRSNVFRQMOQEFQ	240
Qy	241	TFQSHFISDTLTETPFYFFPAFSKEPMTKADLEQOCWDIPNFFQFCNFSVSVSETIT	300

Db 241 TFSHLSIDTLTEPFFAFSKEPMTKADLEQCWDLPNFFQLFCNFSVIYSVSETIT 300  
QY 301 KMLKAIEDLPKQDAPDHGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQCKQAHLS 360  
Db 301 KMLKAIEDLPKQDAPDHGGLISKMLPGQDRGLCGELDONLSRCFKFHEKQCKQAHLS 360  
QY 361 DCPDVPALHTELDEARLVNVSNOQYQGIQIMTKRKHLEDATAYLVKMRGQFGWVSELANO 420  
Db 361 DCPDVPALHTELDEARLVNVSNOQYQGIQIMTKRKHLEDATAYLVKMRGQFGWVSELANO 420  
QY 421 APTEIIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLESASSNFIGYV 480  
Db 421 APTEIIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLESASSNFIGYV 480  
QY 481 VAKALQHFEHFKTW 495  
Db 481 VAKALQHFEHFKTW 495

RESULT 2  
US-09-268-992-64  
; Sequence 64, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 64  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-64

Query Match 93.5%; Score 2460; DB 4; Length 466;  
Best Local Similarity 99.8%; Pred. No. 1e-228;  
Matches 465; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 30 MKPPLLVFVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADEEVKKALTGKQMK 89  
Db 1 MKPPLLVFVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADEEVKKALTGKQMK 60  
QY 90 IMMERKEHTNLMSTLKKREEKQBALKLLNEVQEHLEERLCRESLADSWGECRSL 149  
Db 61 IMMERKEHTNLMSTLKKREEKQBALKLLNEVQEHLEERLCRESLADSWGECRSL 120  
QY 150 ENNCMIYITTCQPSWSVKNKTERFFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOM 209  
Db 121 ENNCMIYITTCQPSWSVKNKTERFFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOM 180  
QY 210 EDVFSOLTVDVNSLNRNFVRQMOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269  
Db 181 EDVFSOLTVDVNSLNRNFVRQMOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 240  
QY 270 DLEQCWDIPNFQFCNFSVSIYSVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 329  
Db 241 DLEQCWDIPNFQFCNFSVSIYSVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 300

QY 330 DRGLCGELDONLSRCKFKHEKQCKQAHLSSEDCPDVPALHTELDEARLVNVSNOQYQGI 389  
Db 301 DRGLCGELDONLSRCKFKHEKQCKQAHLSSEDCPDVPALHTELDEARLVNVSNOQYQGI 360  
QY 390 LQWTRKHLEDATAYLVKMRGQFGWSELANOQAPETELIIFNSIQVVPRIHEGNISKQDETMT 449  
Db 361 LQWTRKHLEDATAYLVKMRGQFGWSELANOQAPETELIIFNSIQVVPRIHEGNISKQDETMT 420  
QY 450 MTDLSILPSSNFTLKIPLESASSNFIGYVYVAKALQHFEHFKTW 495  
Db 421 MTDLSILPSSNFTLKIPLESASSNFIGYVYVAKALQHFEHFKTW 466  
RESULT 3  
US-09-268-992-51  
; Sequence 51, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 51  
; LENGTH: 446  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-51

Query Match 89.1%; Score 2346; DB 4; Length 446;  
Best Local Similarity 99.8%; Pred. No. 9.2e-218;  
Matches 445; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 50 APTWKDKTAISENLKSFSEVGEIDADEEVKKALTGKQMKIMMERKEHTNLMSTLKK 109  
Db 1 APTWKDKTAISENLKSFSEVGEIDADEEVKKALTGKQMKIMMERKEHTNLMSTLKK 60  
QY 110 REEQBALKLLNEVQEHLEERLCRESLADSWGECRSLNENCMRIYITTCQPSWSSVKN 169  
Db 61 REEQBALKLLNEVQEHLEERLCRESLADSWGECRSLNENCMRIYITTCQPSWSSVKN 120  
QY 170 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOMEDVFSOLTVDVNSLNRNSEN 229  
Db 121 KIERFRKIYQFLFPFHEDNEKDLPISEKLIKEDAQLTOMEDVFSOLTVDVNSLNRNSEN 180  
QY 230 VFRQMOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEQCWDIPNFQFCNFSV 289  
Db 181 VFRQMOEEDQTFQSHFISDTDLTEPYFFPAFSKEPMTKADLEQCWDIPNFQFCNFSV 240  
QY 290 SIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDONLSRCKFKHE 349  
Db 241 SIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDRGLCGELDONLSRCKFKHE 300  
QY 350 KQCKQAHLSSEDCPDVPALHTELDEARLVNVSNOQYQGILOWTRKHLEDATAYLVKMRG 409  
Db 301 KQCKQAHLSSEDCPDVPALHTELDEARLVNVSNOQYQGILOWTRKHLEDATAYLVKMRG 360  
QY 410 QFGWVSELANOQAPETELIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLEE 469  
Db 361 QFGWVSELANOQAPETELIIFNSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLEE 420





; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 39  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Cavia sp.  
US-09-268-992-39

Query Match 62.1%; Score 1635; DB 4; Length 466;  
Best Local Similarity 66.5%; Pred. No. 3.4e-149;  
Matches 311; Conservative 64; Mismatches 89; Indels 4; Gaps 3;  
  
QY 30 MKPPLLVTIVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADVEVKKALTGKQMK 89  
DB 1 MKPLPLMFPVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDVGEVKIALIGIKQMK 60  
  
QY 90 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEQHLNEERLRESLADSGWECRSL 149  
DB 61 IMMERREEHSKLMKTLKKCKEEKQKALKLNVEHLEHEESLCOVSLADSWDECRACL 120  
  
QY 150 ENNCRIYTCOPSSWSSVKNKIERFERKIQYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 209  
DB 121 ESNCRFDFTTCOPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 179  
  
QY 210 EDVFSQLTVDVNSLFRNFVRQMOQEFQDTQSFHIFSDTLTEPYFPFAPFSKEPMTKA 269  
DB 180 EHVSQLSADVTSLFNRSLYVFKLRREFDQAFQSYFTSGTDVTEFPFSLSKEPAYRA 239  
  
QY 270 DLEQCWDIPNFQFCNFSVYSEVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPQ 329  
DB 240 DAEPSSWAIPNVFOLLNLSFSVYQSVSEKLITTLRATEDPPKQDKDSNOGGPISKILPEQ 299  
  
QY 330 DRGLCGELDONLSRCFKFHEKQKQAHLSDECPCDVPALHTELDRAIRLVNYSNOQYGOI 389  
DB 300 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPYELRENEALRLVSRNQQYDQV 359  
  
QY 390 LQMTKRHLEDATYLVKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447  
DB 360 VQMTQYHLEDITLLMEKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 419  
  
QY 448 TMTDLSILPSSNFTLKIPLSEASNSNFIQYVYVAKALQHFKEHFKTW 495  
DB 420 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 466  
  
RESULT 7  
US-09-268-992-41  
; Sequence 41, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992

; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 41  
; LENGTH: 450  
; TYPE: PRT  
; ORGANISM: Cavia sp.  
US-09-268-992-41

Query Match 59.0%; Score 1554; DB 4; Length 450;  
Best Local Similarity 63.7%; Pred. No. 2e-141;  
Matches 298; Conservative 63; Mismatches 87; Indels 20; Gaps 3;  
  
QY 30 MKPPLLVTIVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDADVEVKKALTGKQMK 89  
DB 1 MKPLPLMFPVCLLWLDKSHCAPTKDKTAISENLKSFSEVGEIDVGEVKIALIGIKQMK 60  
  
QY 90 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEQHLNEERLRESLADSGWECRSL 149  
DB 61 IMMERREEHSKLMKTLKKCKEEKQKALKLNVEHLEHEESLCOVSLADSWDECRACL 120  
  
QY 150 ENNCRIYTCOPSSWSSVKNKIERFERKIQYQFLFPFHEDNEKDLPISEKLIKDAQLTOM 209  
DB 121 ESNCRFDFTTCOPAWSSVKNM-----ENDRSGPVSKGVTEEDAQVSHI 163  
  
QY 210 EDVFSQLTVDVNSLFRNFVRQMOQEFQDTQSFHIFSDTLTEPYFPFAPFSKEPMTKA 269  
DB 164 EHVSQLSADVTSLFNRSLYVFKLRREFDQAFQSYFTSGTDVTEFPFSLSKEPAYRA 223  
  
QY 270 DLEQCWDIPNFQFCNFSVYSEVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPQ 329  
DB 224 DAEPSSWAIPNVFOLLNLSFSVYQSVSEKLITTLRATEDPPKQDKDSNOGGPISKILPEQ 283  
  
QY 330 DRGLCGELDONLSRCFKFHEKQKQAHLSDECPCDVPALHTELDRAIRLVNYSNOQYGOI 389  
DB 284 DRGSDGKLGONLSDCNVFRKRCQKQDYLSDDCPNVPYELRENEALRLVSRNQQYDQV 343  
  
QY 390 LQMTKRHLEDATYLVKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 447  
DB 344 VQMTQYHLEDITLLMEKMRQFGWVSELANQAPETEIFNSIQVVPRI--HEGNISKQDE 403  
  
QY 448 TMTDLSILPSSNFTLKIPLSEASNSNFIQYVYVAKALQHFKEHFKTW 495  
DB 404 TVVPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFKTW 450  
  
RESULT 8  
US-09-268-992-43  
; Sequence 43, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05

; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 43  
; LENGTH: 374  
; TYPE: PR1  
; ORGANISM: Cavia sp.  
US-09-268-992-43

Query Match 48.5%; Score 1277; DB 4; Length 374;  
Best Local Similarity 54.5%; Pred. No. 7.7e-115;  
Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;  
Qy 30 MKPPLLFIIVCLLWKDCHCAPTKWKTATSENLKSFSEVGEIDADEVKKALTGKQMK 89  
Db 1 MKPLLPFPVCLLWKDCHCAPTKWKTATSENLKSFSEVGEIDADEVKKALTGKQMK 60  
Qy 90 IMMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149  
Db 61 IMMERREEHSHKMLTKLKKCKEEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 120  
Qy 150 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEKDAQLTQM 209  
Db 121 ESNCRFDTTCCPAWSSVKN----- 140  
Qy 210 EDVFSQTLVDVNSLNFNSFNVRQMQEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269  
Db 141 -----NEPAYRA 147  
Qy 270 DLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKATEDLPKQKADPHGGLSKMLPGQ 329  
Db 148 DAEPWALPNVQLLNLNLSFVSQVSEKLTTLRATEDPPKQKDSNOGPGISKILPEQ 207  
Qy 330 DRGLCGELDONLRCFKFHEKCKQKQHLSEDPCDVPALHTELDEAIRLVNVSNOQYQGI 389  
Db 208 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPYELRELNEALRLVRSRNOQYDQV 267  
Qy 390 LQMTKHELDYAVLEKMRGQFGWVSELANOAPETIIFNSIOVVPRI--HEGNISKODE 447  
Db 268 VQMTQYHLEDTLLMEKMRQFGWVSELAYQSPGAEDIFPNVKKVLAHQHFKETW 374  
Qy 448 TMMTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHFKEHFTW 495  
Db 328 TVWPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFTW 374

RESULT 9  
US-09-268-992-45  
; Sequence 45, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 45  
; LENGTH: 373  
; TYPE: PR1

; ORGANISM: Cavia sp.  
US-09-268-992-45

Query Match 48.3%; Score 1272.5; DB 4; Length 373;  
Best Local Similarity 54.5%; Pred. No. 2.1e-114;  
Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

Qy 30 MKPPLLFIIVCLLWKDCHCAPTKWKTATSENLKSFSEVGEIDADEVKKALTGKQMK 89  
Db 1 MKPLLPFPVCLLWKDCHCAPTKWKTATSENLKSFSEVGEIDADEVKKALTGKQMK 60  
Qy 90 IMMERKEHTNLMSTLKKCKREEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 149  
Db 61 IMMERREEHSHKMLTKLKKCKEEKQKALKLLNEVQHELEERLCRESLADSWGECRSL 120  
Qy 150 ENNCRIYTTCCPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLEKDAQLTQM 209  
Db 121 ESNCRFDTTCCPAWSSVKN----- 140  
Qy 210 EDVFSQTLVDVNSLNFNSFNVRQMQEFDQTFQSHFISDTDLTEPYFFPAFSKEPMTKA 269  
Db 141 -----MPAYRA 146  
Qy 270 DLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKATEDLPKQKADPHGGLSKMLPGQ 329  
Db 147 DAEPWALPNVQLLNLNLSFVSQVSEKLTTLRATEDPPKQKDSNOGPGISKILPEQ 206  
Qy 330 DRGLCGELDONLRCFKFHEKCKQKQHLSEDPCDVPALHTELDEAIRLVNVSNOQYQGI 389  
Db 207 DRGSDGKLGONLSDCVNFRKRCQKQDYLSDDCPNVPYELRELNEALRLVRSRNOQYDQV 266  
Qy 390 LQMTKHELDYAVLEKMRGQFGWVSELANOAPETIIFNSIOVVPRI--HEGNISKODE 447  
Db 267 VQMTQYHLEDTLLMEKMRQFGWVSELAYQSPGAEDIFPNVKKVLAHQHFKETW 373  
Qy 448 TMMTDLSILPSSNFTLKIPLESSESSNFIYGVYVAKALQHFKEHFTW 495  
Db 327 TVWPS-SLLPSSNFTLSSPLEKSAGNANFIDHVVKEVQLQHFKEHFTW 373

RESULT 10  
US-09-268-992-67  
; Sequence 67, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 67  
; LENGTH: 208  
; TYPE: PR1  
; ORGANISM: Homo sapiens  
US-09-268-992-67

Query Match 35.0%; Score 921; DB 4; Length 208;  
Best Local Similarity 99.4%; Pred. No. 6.5e-81;  
Matches 170; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy	1	MK1AEKNEGPSRWQHLHGWDIANNNGNNKPPLLVFI	CLLWLKDSHCAPTWKDKTALS	60
Db	33	MK1AEKNEGPSRWQHLHGWDIANNNGNNKPPLLVFI	CLLWLKDSHCAPTWKDKTALS	92
Qy	61	ENLKSSEVGEIDADAEVVKALGTGIQOMKIMMERKE	HEHTNLMSTLKKCREKQBAKLL	120
Db	93	ENLKSSEVGEIDADAEVVKALGTGIQOMKIMMERKE	HEHTNLMSTLKKCREKQBAKLL	152
Qy	121	NEVQEHLEBERLCRSLADSGECRSCLENNCMRIYTT	QCPSWSGVNKKI	171
Db	153	NEVQEHLEBERLCRSLADSGECRSCLENNCMRIYTT	QCPSWSGVNKKI	203

```

RESULT 11
US-08-685-576-1
; Sequence 1, Application US/08685576
; Patent No. 5906819
; GENERAL INFORMATION:
; APPLICANT: Kaibuchi, Kozo
; APPLICANT: Iwamatsu, Akinori
; APPLICANT: Nakano, Takeshi
; APPLICANT: Ito, Masaaki
; APPLICANT: Takahashi, No. 5906819aaki
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE
; NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W., Suite 500
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/685,576
; FILING DATE: 24-JUL-1996

```

Query Match	5.7%;	Score 149;	DB 2;	Length 1388;
Best Local Similarity	16.7%;	Pred. No. 3e-05;		
Matches 95;	Conservative 96;	Mismatches 166;		
		Indels 212;	Gaps 20;	
QY	54	KDKTAISENLKSPFSEVEIGDADEVKALGCIQMKI-----	MMERKKEHTNLMS	104

Db	780	QKQDV	LNEDV	RLNLT	---	LKIE	QKQ	RLCTQ	NDL	KMQ	TQV	NLT	KM	SE	KQ	LN	NHLL	835
Qy	105	TLK	---	---	---	KCRE	KQ	EAL	KL	NVE	Q	HE	LE	---	---	---	---	135
Db	836	EMK	SL	E	KQ	NAEL	KR	QD	AG	Q	M	KEL	Q	DL	E	AQ	---	895
Qy	136	E	---	---	---	SLAD	S	GC	R	SC	LE	N	C	M	R	I	Y	189
Db	896	ELQ	Q	K	Q	E	L	O	D	E	---	---	---	---	---	---	---	942
Qy	190	E	---	---	---	KLIE	K	D	A	O	L	T	O	M	E	D	V	240
Db	943	KIM	LE	I	K	E	M	A	R	H	K	O	E	L	T	D	A	1002
Qy	241	T	F	G	H	F	T	S	D	T	D	L	E	P	E	P	F	300
Db	1003	LSR	---	---	---	LKDE	E	I	S	A	A	A	I	K	A	F	E	1036
Qy	301	K	M	L	K	A	I	B	---	---	---	---	---	---	---	---	---	345
Db	1037	K	L	A	E	I	M	N	K	P	V	R	G	N	D	T	V	1082
Qy	346	K	P	H	E	K	C	O	K	A	H	L	S	E	D	C	P	398
Db	1083	K	Y	O	K	E	L	N	E	M	O	A	I	A	E	---	---	1138
Qy	399	---	---	---	---	DTAY	L	V	E	K	M	R	G	---	---	---	---	417
Db	1139	P	G	T	E	A	D	D	G	F	E	S	R	L	E	G	W	1198
Qy	418	---	---	---	---	ANQ	A	P	E	I	F	I	N	S	I	Q	V	462
Db	1199	V	L	D	I	D	K	L	F	H	R	V	P	T	O	D	V	1243
Qy	463	L	K	T	P	L	E	S	A	E	S	S	N	F	I	---	---	488
Db	1244	---	---	---	---	F	P	V	E	G	K	S	N	Y	I	C	H	---

RESULT 12  
US-08-685-576-4  
; Sequence 4, Application US/08685576  
; Patent No. 5906819  
; GENERAL INFORMATION:  
; APPLICANT: Kaibuchi, Koza  
; APPLICANT: Iwamatsu, Akihiro  
; APPLICANT: Nakano, Takeshi  
; APPLICANT: Ito, Masaaki  
; APPLICANT: Takahashi, No. 5906819usaki  
; TITLE OF INVENTION: RHO TARGET PROTEIN RHO-KINASE  
; NUMBER OF SEQUENCES: 16  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Foley & Lardner  
; STREET: 3000 K Street, N.W., Suite 500  
; CITY: Washington  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20007-5109  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/685,576  
; FILING DATE: 24-JUL-1996  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 7-325129  
; FILING DATE: 20-NOV-1995  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 8-17150

```

: FILING DATE: 05-JAN-1996
:
: PRIORITY APPLICATION DATA:
:
: APPLICATION NUMBER: JP 8-131206
: FILING DATE: 26-APR-1996
: ATTORNEY/AGENT INFORMATION:
:
: NAME: Bent, Stephen A.
:
: REGISTRATION NUMBER: 29,768
: REFERENCE/DOCKET NUMBER: 16887/843
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)672-5300
: TELEFAX: (202)672-5399
:
: TELEX: 904136
:
: INFORMATION FOR SFO ID NO: 4:
:
: SEQUENCE CHARACTERISTICS:
:
: LENGTH: 1388 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
:
: MOLECULE TYPE: protein
:
: US-08-685-576-4

```

Query Match	5.5%	Score 145;	DB 2;	Length 1388;
Best Local Similarity	17.0%;	Pred. No. 7.4e-05;		
Matches 95;	Conservative	96;	Mismatches 176;	Indels 192;
				Gaps 20;

QY	54	KDTAISENLSKFSEVEIDADEVKALITGOMKI-----MWERKEHETNLMS	104
Db	780	KORDVLINEDVRNLT----LKIEOTEKRCLTQNDLMQMOTQQVNTLKMSEKQLQENNHLM	835
QY	105	TLK-----KCREEKOEALKLLNEVOHLEPEE---RLCRESIADSWGEC-----	145
Db	836	EMXHNLEKQNAUERKERQRDAGOMKELODOLAEQFSTLYQTQVRELKECEECKTLKG	895
QY	146	---RSLENNCRMIYTTCPSPSSVNKNTERFRKIYOFLPFFHEDNE--KDLPIS---	197
Db	896	ELQOKQEQLQDERSDLAQALETILTKADSEQULARSIAEQYSDEKEIKMLEITEKMA	955
QY	198	---KLIEDAQLTQMEDVSFLTVDVNSLFNRS---FNVPROMOQEFDOTFSHFISDT	250
Db	956	RHKQELTERDATIASLEETNRTLTDVANLANEKELNNKLKDVOEQLSR-----LKDE	1009
QY	251	DLTEPYFFPAFSKEPMTKADLECWDIPNFQOLFNCFSVIYESVETITMLKAIE---	307
Db	1010	EISAAAIKAQFEKOLLUTETLK-----TOAVNKLAETIMNRKE	1046
QY	308	-----DLPGDKAPDHGGILSKMLPGDRGLCGEL----DONLSRCFKFHCKQCQ	355
Db	1047	PVKRGNDTVRRKE-----ENRKLHWELKSEREKITQOMIKYQKELNMQ	1092
QY	356	AHLSEDCPDVALHTELDRAIRLVNVNQYQGIILOMTRKHLE-----DTAY	402
Db	1093	AQIAEE---SOIRIELOWLTDSKSDIEQLRSQALHIGLDSSSGSGPCDAENDGF	1148
QY	403	LVEKHMG-----QFGHVSEL-----A	418
Db	1149	PSRLEGWLSLPVRNNTKFGWVKYIVSVSSKKILFYDSEQKESPNMYVLDIDKLFHV	1208
QY	419	NOAPETEIIINSIQVVPRI-----HEGNISKODETMWTDLSILPSSNFLTKIPLESAE	472
Db	1209	RPWTQDVYRADAKELPRIFQLLYANEGESKQE-----FPVEPVGE	1251
QY	473	SNFTI---GYVVAKALQH F	488
Db	1252	KSNYICHKGHEFIPTLYHF	1270

RESULT 13  
US-08-533-306A-6  
; Sequence 6, Application US/085333306A  
; Patent No. 5837457  
; GENERAL INFORMATION:  
; APPLICANT: Liu, Pu  
; APPLICANT: Collins, Francis S.

```

1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841
842
843
844
845
846
847
848
849
850
851
852
853
854
855
856
857
858
859
860
861
862
863
864
865
866
867
868
869
870
871
872
873
874
875
876
877
878
879
880
881
882
883
884
885
886
887
888
889
890
891
892
893
894
895
896
897
898
899
900
901
902
903
904
905
906
907
908
909
910
911
912
913
914
915
916
917
918
919
920
921
922
923
924
925
926
927
928
929
930
931
932
933
934
935
936
937
938
939
940
941
942
943
944
945
946
947
948
949
950
951
952
953
954
955
956
957
958
959
960
961
962
963
964
965
966
967
968
969
970
971
972
973
974
975
976
977
978
979
980
981
982
983
984
985
986
987
988
989
990
991
992
993
994
995
996
997
998
999
1000
1001
1002
1003
1004
1005
1006
1007
1008
1009
1010
1011
1012
1013
1014
1015
1016
1017
1018
1019
1020
1021
1022
1023
1024
1025
1026
1027
1028
1029
1030
1031
1032
1033
1034
1035
1036
1037
1038
1039
1040

```

Query Match 5.0%; Score 131; DB 2; Length 816;  
Best Local Similarity 19.0%; Pred. NO. 0.00073;  
Matches 96; Conservative 81; Mismatches 177; Indels 1

QY	4	KAEKNEGSPRSWQOLHWGDJANNNGN---MKPPLLVFIVICLLMLKDSHCAPTWKDKTAISE	61
Db	73	QGEQROTSPR---EYVDLEREAGKVYLKAPMILNGVCVW---KGWID-----	114
QY	62	NLKSFSEVSGIDADEE---VKKALTGIKQMKIMMERKEKEHTNLM-----STLKK	108
Db	115	-LQRLDGMGCLFEDEERAQEDALAAQAAFEABRRRTREFDRDRSHREMEMEVEVSVTGM	173
QY	109	CREEKQEAULK-----LNEVEQHLEEEERLC-----RESLADSWGBC	145
Db	174	LNEAEGKAIKLAKDVASLSQLODQTQELLQEETROKLVNSTKLROLAEERNSIQLOLDSE	233
QY	146	RSCLENNCMIYTCOPSSWSSVKNKTERFRKFIQPLFPFHEDNEKDLP-----ISEKLI	200
Db	234	MEA-KONLERHISTLNIQSDSKKKLODFASTVEAL-----BEGKRKFQKEIENITQQYE	287
QY	201	EKDAQLTOMEDVFSQLTVDVNSLF-----NRSFNFRQMQEFDQ-----TFQSHFIS	248
Db	288	EKAAAYDKLEKTNRLQEQELDDLVDLDNQROLVSNLEKKQKFDOLLAEEKNISKKYAD	347
QY	249	DTDLTEPYFPFAPSKEPMFKADLEQCWDIPNRFQLPFCNFSVIYSSVSTIYTKMLKAIED	308
Db	348	ERDRAE-----AGAREKETKA-----LSLARALEEALEAKEE	379
QY	309	LPKQDKAPDHGGILSKMLPQDQRGLCGELDQNLRSRCFKPHEKCOKCOAHLSEDCPDVPAL	368
Db	380	LER-----TNKMLAEEMEDLVSSKDDYGVKNVHELEKSKRALETQMEEMKTLQEL	429
QY	369	HTEL-----DEARL-VNVSNOOYGGIQLQWTRKHLEDYAVLVEKMRQGFQGWSELANQAPE	423
Db	430	EDLSQASEDAKURLEVN-----MOALKQGF-----ERDLOARD	462





**THIS PAGE BLANK (USPTO)**

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:52:07 : Search time 17.57 Seconds  
(without alignments)  
1090.847 Million cell updates/sec

Title: US-09-722-544A-2MOD

Perfect score: 2632

Sequence: 1 MKIAEKNEGFSRWQLHW.....FIGYVAKALQHFKEHFKTW 495

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	434	16.5	449	1	CLUS_HUMAN
2	418	15.9	451	1	CLUS_COTJA
3	413.5	15.7	447	1	CLUS_RABIT
4	410.5	15.6	445	1	CLUS_CANFA
5	403	15.3	448	1	CLUS_MOUSE
6	400.5	15.2	439	1	CLUS_BOVIN
7	400.5	15.2	446	1	CLUS_PIG
8	380	14.4	447	1	CLUS_RAT
9	168	6.4	191	1	CLUS_MESAU
10	151	5.7	1526	1	MYS2-SCHPO
11	145.5	5.5	594	1	CIK1_YEAST
12	145.5	5.5	1938	1	MYHD_HUMAN
13	141.5	5.4	1005	1	RA50_METJA
14	141.5	5.4	3433	1	UTRO_HUMAN
15	139.5	5.3	2663	1	CENE_HUMAN
16	138	5.2	5430	1	ACFT_HUMAN
17	137	5.2	1940	1	MYH3_HUMAN
18	137	5.2	2230	1	GOG4_HUMAN
19	135.5	5.1	1679	1	YIO9_YEAST
20	135	5.1	1060	1	EG51_XENLA
21	133.5	5.1	1957	1	YD86_SCHPO
22	132.5	5.0	852	1	RA50_THEMA
23	131.5	5.0	1978	1	MYHB_CHICK
24	130.5	5.0	971	1	Y228_BORBU
25	130.5	5.0	1940	1	MYH3_RAT
26	130	4.9	793	1	KATA_ARATH
27	129.5	4.9	1427	1	REST_HUMAN
28	129.5	4.9	2077	1	TEGU_HSV6U
29	129	4.9	539	1	MYS3_HYDAT
30	129	4.9	1939	1	MYH1_HUMAN
31	128	4.9	1744	1	TANA_XENLA
32	128	4.9	1853	1	MY5A_MOUSE
33	128	4.9	2704	1	BPAL_HUMAN

34	128	4.9	2869	1	BBP1_PLAVB
35	127.5	4.8	722	1	MFPI_TOBAC
36	127.5	4.8	1875	1	MLPI_YEAST
37	127	4.8	1189	1	SCII_CHICK
38	126	4.8	1941	1	MYH2_HUMAN
39	125.5	4.8	571	1	YHY4_YEAST
40	125.5	4.8	978	1	RA50_AQUAE
41	125.5	4.8	1067	1	EG52_XENLA
42	125.3	4.8	1966	1	MYSB_CAEEL
43	125.5	4.8	1969	1	MYSB_CAEEL
44	125.5	4.8	2198	1	YLJ2_CAEEL
45	125.5	4.8	3210	1	CENF_HUMAN

ALIGNMENTS

RESULT	1
CLUS_HUMAN	
ID	CLUS_HUMAN STANDARD; PRT; 449 AA.
AC	P10909; P11380; P11381;
DT	01-JUL-1989 (Rel. 11, Created)
DT	01-JUL-1989 (Rel. 11, Last sequence update)
DT	01-MAR-2002 (Rel. 41, Last annotation update)
DE	Clusterin precursor (Complement-associated protein SP-40,40)
DE	(Complement cytotoxicity inhibitor) (CLI) (NAL and NA2) (Apollipoprotein J) (Apo-J) (TRPM-2).
GN	CLU.
OS	Homo sapiens (Human).
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX	NCBI_TaxID=9606;
RN	[1]
RP	SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC	TISSUE=Liver;
RX	MEDLINE=89251601; PubMed=2721499;
RA	Kirschbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,
RA	Hudson P., Walker I.D.;
RT	"Molecular cloning and characterization of the novel, human
RT	complement-associated protein, SP-40,40: a link between the
RT	complement and reproductive systems.";
RL	EMBO J. 8:711-718(1989).
RN	[2]
RP	SEQUENCE OF 2-449 FROM N.A.
RX	MEDLINE=89386692; PubMed=2780565;
RA	Tenne D.E., Tschoop J.;
RT	"Molecular structure and functional characterization of a human
RT	complement cytotoxicity inhibitor found in blood and seminal plasma:
RT	identity to sulfated glycoprotein 2, a constituent of rat testis
RT	fluid.";
RL	Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).
RN	[3]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=90344779; PubMed=1974459;
RA	de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;
RT	"Apollipoprotein J: structure and tissue distribution.";
RL	Biochemistry 29:5380-5389(1990).
RN	[4]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=94237156; PubMed=8181474;
RA	Wong P., Taillefer D., Lakin J., Pineault J., Chader G.,
RA	Tenniswood M.;
RT	"Molecular characterization of human TRPM-2/clusterin, a gene
RT	associated with sperm maturation, apoptosis and neurodegeneration.";
RL	Eur. J. Biochem. 221:917-925(1994).
RN	[5]
RP	SEQUENCE FROM N.A.
RX	MEDLINE=93186813; PubMed=7680346;
RA	Pineault J.M., Tenniswood M.;
RT	"Genomic organization and expression of the rat TRPM-2 (clusterin)
RT	gene, a gene implicated in apoptosis.";
RL	J. Biol. Chem. 268:5021-5031(1993).
RN	[6]

RP SEQUENCE FROM N.A.  
RA Blechschmidt K., Rosenthal A.;  
RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 61-449 FROM N.A.  
RX MEDLINE-92020896; PubMed-1924317;  
RA Dank M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,  
RA Quirion R., Suh M.;  
RT "Human gliomas and epileptic foci express high levels of a mRNA  
RT related to rat testicular sulfated glycoprotein 2, a purported marker  
RT of cell death.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).  
RN [8]  
RP SEQUENCE OF 140-449 FROM N.A.  
RC TISSUE=Liver;  
RA Glew M.D., Kirsbaum L., Bozas S.E., Walker I.D.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP SEQUENCE OF 23-33 AND 228-240.  
RX MEDLINE-91265608; PubMed-3154963;  
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,  
RA Hochstrasser D.F., Pometta D., Merrill C.R.;  
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:  
RT identification and partial sequencing";  
RL Appl. Theor. Electrophor. 1:73-76(1988).  
RN [10]  
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.  
RX MEDLINE-91230083; PubMed-1903064;  
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,  
RA Hochstrasser D.F.;  
RT "Characterization of a human high density lipoprotein-associated  
RT protein, NAI/NA2. Identity with SP-40, an inhibitor of complement-  
RT mediated cytotoxicity.";  
RL Arterioscler. Thromb. 11:645-652(1991).  
RN [11]  
RP SEQUENCE OF 23-52 AND 228-257.  
RX MEDLINE-90354412; PubMed-2387851;  
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,  
RA Wetterau J.R., Busch S.J., Harmony J.A.K.;  
RT "Purification and characterization of apolipoprotein J";  
RL J. Biol. Chem. 265:14292-14297(1990).  
RN [12]  
RP SEQUENCE OF 23-37 AND 228-242.  
RX MEDLINE-90097955; PubMed-2601725;  
RA Choi N.H., Mazda T., Tomita M.;  
RT "A serum protein Sp40,40 modulates the formation of membrane attack  
RT complex of complement on erythrocytes.";  
RL Mol. Immunol. 26:835-840(1989).  
RN [13]  
RP SEQUENCE OF 23-41 AND 228-246.  
RX MEDLINE-93319521; PubMed-8328966;  
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,  
RA Wisniewski T., Frangione B.;  
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is  
RT complexed to SP-40,40 (apolipoprotein J), an inhibitor of the  
RT complement membrane-attack complex.";  
RL Biochem. J. 293:27-30(1993).  
RN [14]  
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE-92201397; PubMed-1551440;  
RA Kirsbaum L., Bozas S.E., Walker I.D.;  
RT "SP-40,40, a protein involved in the control of the complement  
RT pathway, possesses a unique array of disulphide bridges.";  
RL FEBS Lett. 297:70-76(1992).  
RN [15]  
RP DISULFIDE BONDS.  
RX MEDLINE-93147076; PubMed-1491011;  
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;  
RT "Identification of the disulfide bonds in human plasma protein  
RT SP-40,40 (apolipoprotein-J).";  
RL J. Biochem. 112:557-561(1992).  
RN [16]  
RP CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE-97477902; PubMed-9336835;  
RA Kapron J.T., Hilliard G.M., Laking J.N., Tenniswood M.P., West K.A.,  
RA Carr S.A., Crabb J.W.;  
RT "Identification and characterization of glycosylation sites in human  
RT serum clusterin.";  
RL Protein Sci. 6:2120-2133(1997).  
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH (APOPTOSIS).  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X14723; CAA32847.1; -;  
DR EMBL; M25915; AAA35692.1; ALT\_INIT.  
DR EMBL; J02908; AAA51765.1; ALT\_INIT.  
DR EMBL; M74816; AAA60321.1; -;  
DR EMBL; M64722; AAA606508.1; -;  
DR EMBL; L00974; AAA606507.1; -;  
DR EMBL; AF311103; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M63376; AAB06507.1; -;  
DR EMBL; M63377; AAB06507.1; JOINED.  
DR EMBL; M63378; AAB06507.1; JOINED.  
DR EMBL; A21577; CAA01560.1; -;  
DR PIR; A35833; A35833.  
DR PIR; A37816; A37816.  
DR PIR; B37816; B37816.  
DR PIR; A41386; A41386.  
DR PIR; PLO135; PLO135.  
DR PIR; PLO136; PLO136.  
DR PIR; S04662; S04662.  
DR PIR; S07433; S07433.  
DR PIR; S09339; S09339.  
DR PIR; S34056; S34056.  
DR PIR; S43646; S43646.  
DR SWISS-2DPAGE; P10909; HUMAN.  
DR MIM; 185430; -;  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; Cla; 1.  
DR SMART; SM00030; Cld; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;  
KW Signal.  
FT SIGNAL. 1 22  
FT CHAIN 23 449  
FT CHAIN 23 227  
FT CHAIN 228 449  
FT CHAIN 102 313  
FT DISULFID 113 305  
FT DISULFID 116 302  
FT DISULFID 121 295  
FT DISULFID 129 285  
FT CARBOHYD 86 86  
FT CARBOHYD 103 103  
FT CARBOHYD 145 145  
FT CARBOHYD 291 291  
FT CARBOHYD 354 354  
FT CARBOHYD 374 374  
FT CARBOHYD 28 28  
FT CONFLICT 47 47  
FT CONFLICT 52 52  
FT CONFLICT 305 305  
FT CLUSTERIN.  
FT BETA-CHAIN.  
FT ALPHA-CHAIN.  
FT INTERCHAIN.  
FT INTERCHAIN.  
FT INTERCHAIN.  
FT INTERCHAIN.  
FT INTERCHAIN.  
FT N-LINKED (GLCNAC. . .).  
FT N-LINKED (GLCNAC. . .).  
FT N-LINKED (GLCNAC. . .).  
FT N-LINKED (GLCNAC. . .).  
FT N-LINKED (GLCNAC. . .).  
FT D -> S (IN REF. 9 AND 10).  
FT G -> H (IN REF. 11).  
FT G -> Q (IN REF. 11).  
FT C -> M (IN REF. 10).

Query Match 16.5%; Score 434; DB 1; Length 449;  
 Best Local Similarity 25.8%; Pred. No. 2e-19;  
 Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 30 MKPPLLVIVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEGEIDAEDEVKKALT 83  
 DB 1 MKKTLVLLVGLLL-----TWESQVLGDQTVSDNELQMSNQSKYVKNQIAVAVN 51

QY 84 GIKOMKIMMERKEHTNLMTLKKCKREKQKALKNLNEVQHELEERLCRSLADSWG 143  
 DB 52 GVQIKTLIBKNEERTKLSLNEEAKKEDALNETRESETKLKELPGVCNETMMALWE 111

QY 144 ECRSCLENNKMRYT-TCQPSWSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLEIK 202  
 DB 112 ECRPCLQKCMKFVYRCRSGSLVGLQLEELNQSFPYFWNGDR-----IDSLLEN 165

QY 203 DAQLTQMEDV-----FSQLTVDVNSLNRNFVFRQMOQEDQTFQSHFIS-DTDLTPEYF 257  
 DB 166 DROOTHMLDVMDHFSRASSIIDLEFQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218

QY 258 FPAFSKEPMTKADLE-OCWDIPNFFOLFNCFSVSIYESVSETITKMLKAIEDLPKQDKAP 316  
 DB 219 F--FPKSRIVRSLMPPSPYPLNFHAFMQPFLMIHEA-----QQAMDTHFHSAPAF 267

QY 317 DHGGLISKMLPGQDGLGELQNLSCFKFHEKCKQCAHLSDEC-----PDVPALHTEL 372  
 DB 268 QHPTEPIREGDDRTVCREIRHNSTGCLRMKQCDCKREILSDVCSSTNPNPQAKLREL 327

QY 373 DEAIRLVNNSQYQGLQWTRKHLEDTAYLVKMRGQFGWSELANOAPETEIFNSIQ 432  
 DB 328 DESLOAERLTRKYNELLKQYKMLNTSLLEQNEQFNQVNSRLANLTQGEDQYLRVT 387

QY 433 VVPRIHGNISKODETMMTDLSTLPSNFTLKIPLESSESSNFIQVYVAKALQHF-KEH 491  
 DB 388 TVAS-HTSDSDVSGVTEVVVKLFDSPTITVTPVEVSRNKPFRMETVAEKALQYRKKH 446

RESULT 2  
 CLUS\_COTJA  
 ID CLUS\_COTJA STANDARD; PRT; 451 AA.

AC P14018;  
 DT 01-JAN-1990 (Rel. 13, Created)  
 DT 01-OCT-1996 (Rel. 34, Last sequence update)  
 DT 15-DEC-1998 (Rel. 37, Last annotation update)  
 DE Clusterin precursor (51.5 kDa protein).  
 GN T64.  
 OS Coturnix coturnix japonica (Japanese quail).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Archosauria; Aves; Neognathae; Galliformes; Phasianinae;  
 OC Coturnix.  
 OX NCBI\_TaxID=93934;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC TISSUE=Neuroretina;  
 RX MEDLINE=89239492; PubMed=2541393;  
 RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;  
 RT "Expression of a novel gene encoding a 51.5 kD precursor protein is  
 induced by different retroviral oncogenes in quail neuroretinal  
 cells.";  
 RL Oncogene Res. 4:127-136(1989).  
 RN [2]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=95262670; PubMed=7744033;  
 RA Michel D., Chatelain G., Herault Y., Brun G.;  
 RT "The expression of the avian clusterin gene can be driven by two  
 alternative promoters with distinct regulatory elements.";  
 RL Eur. J. Biochem. 229:215-223(1995).  
 CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
 CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
 CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
 CC CELL DEATH.  
 CC -!- SUBUNIT: ANTI-PAALLEL DISULFIDE-LINKED HETERODIMER

(BY SIMILARITY).  
 -!- INDUCTION: BY DIFFERENT RETROVIRAL ONCOGENES.  
 -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

CC -----  
 CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its  
 CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (see http://www.isb-sib.ch/announce/  
 CC or send an email to license@sib-sib.ch).

CC -----  
 CC EMBL; X15825; CAA33823.1; -;  
 CC EMBL; X80760; CAA56733.1; -;  
 CC PIR; S07714; S07714; Clusterin.  
 CC InterPro: IPR000753; Clusterin.  
 CC Pfam; PF01093; Clusterin; 1.  
 CC SMART; SM00035; Cla; 1.  
 CC SMART; SM00030; Clb; 1.  
 CC PROSITE; PS00492; CLUSTERIN\_1; 1.  
 CC PROSITE; PS00493; CLUSTERIN\_2; 1.  
 CC Glycoprotein; Signal.  
 CC SIGNAL 1 18 POTENTIAL.  
 CC CHAIN 19 451 CLUSTERIN.  
 CC CHAIN 19 230 BETA-CHAIN (SMALL SUBUNIT) (POTENTIAL).  
 CC CHAIN 231 451 ALPHA-CHAIN (LARGE SUBUNIT) (POTENTIAL).  
 CC DISULFID 98 314 INTERCHAIN (BY SIMILARITY).  
 CC DISULFID 109 306 INTERCHAIN (BY SIMILARITY).  
 CC DISULFID 112 303 INTERCHAIN (BY SIMILARITY).  
 CC DISULFID 117 296 INTERCHAIN (BY SIMILARITY).  
 CC DISULFID 125 286 INTERCHAIN (BY SIMILARITY).  
 CC CARBOHYD 99 99 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 141 141 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 278 278 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 355 355 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CARBOHYD 375 375 N-LINKED (GLCNAC. .) (POTENTIAL).  
 CC CONFLICT 437 437 A -> P (IN REF. 1).  
 CC CONFLICT 445 451 KONNTIE -> SRTTP (IN REF. 1).  
 CC SEQUENCE 451 AA; 51800 MW; 8CF811225C5B22F3 CRC64;

Query Match 15.9%; Score 418; DB 1; Length 451;  
 Best Local Similarity 25.7%; Pred. No. 1.8e-18;  
 Matches 118; Conservative 94; Mismatches 179; Indels 68; Gaps 15;

QY 63 LKSFSEGEIDAEDEVKKALTGKIMMERKEHEHTNLMTLKKCKREKQKALKNLE 122  
 DB 27 LKQLSAAQSYDAEVENAINGVKQMTLMDKTSKEHQAMLHLETKKKEEAVKLALE 86

QY 123 VQHELEERLCRSLADSWGECRSCLENMCMRYT-TCQPSWSVKNKIERFRKIYQF 181  
 DB 87 KEKQLAKEQVNCNETMLSLWECKPCLKHKTCMRVYKMCSSGSLVGLQLEEFNRSPP 146

QY 182 LFPFHEDNEKDLPISEKLEIK-----DAQLTQMEDVFSQLTVDVNSLNRNFVFRQMOQ 236  
 DB 147 SIWNGERIDDLDRQQRERRFEDLEERGLMEDGVEDFQDSTQLYGPAFFPFR---- 202

QY 237 EFDQTFOSHISDTDLTEPY--FFPAFSKEPMTKADLEQCDWDIPN-----FFQLFC 285  
 DB 203 -----TPPFGGFREAF-VPPVQVRVHL-----VPRRLSRELHFFHFFQHPM 240

QY 286 NFSVSIYESVSETITKMLKAIEDLPKQDKADPH--GGLISKMLP-GODRGLCKEDLNLS 342  
 DB 241 HGFHRLFPQLEFMTQHMDG-----CHGAWHEPLGFGFATERNFSDRWVCRIRNSA 294

QY 343 RCFKFEKCKQCAHLSDEC-----PDVPALHTELDEAIRLVNNSQYQGLQWTRKHLE 398  
 DB 295 GCLMRDECEKREILLAVDCSQTDVPVQSQLREQPEDALRLAERTTRYDILLSAFAEML 354

QY 399 DTAYLVKMRGQFGWSELANOAP-----ETEIFNSIQVVPRIHGNISKODETMMT 452  
 DB 355 NTSSLDLQNLQRFQGWVNSRLNLTQGNDFGLQVTTVFSK---TPNLEDP--SAPADTQVT- 408

```
QY 453 LSILPSNFTLKIPLSEAESNFICYVYVAKALQHFKEH 491
DB 409 VQLFSEPLSTVPGDISWDDPRFMEIVAEQALQHYKQN 447

RESULT 3
CLUS_RABIT
ID CLUS_RABIT STANDARD; PRT; 447 AA.
AC Q9XSC5;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE Clusterin precursor (Apolipoprotein J) (Apo-J).
GN CLU.
OS Oryctolagus cuniculus (Rabbit).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
OX NCBI_TaxID=9986;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAPANESE WHITE;
RA Miyata M., Biro S., Kaleda H., Eto H., Orihara K., Kihara T.,
RA Obata H., Matsushita N., Matsuyama T., Tei C.;
RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells
RT by balloon injury.";
RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; AF118852; AAD24461.1; -
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; Cla; 1.
DR SMART; SM00030; Clb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
KW Glycoprotein; plasma; signal.
FT SIGNAL 1 22
FT CHAIN 23 447
FT CHAIN 23 447
FT CHAIN 225 447
FT CHAIN 226 447
FT DISULFID 101 311
FT DISULFID 112 303
FT DISULFID 115 300
FT DISULFID 120 293
FT DISULFID 128 283
FT CARBOHYD 85 85
FT CARBOHYD 102 102
FT CARBOHYD 144 144
FT CARBOHYD 289 289
FT CARBOHYD 326 326
FT CARBOHYD 352 352
FT CARBOHYD 372 372
SQ SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A GRC64;
```

```
Query Match 15.7%; Score 413.5; DB 1;
Best Local Similarity 24.1%; Pred. No. 3.4e-18;
Matches 111; Conservative 93; Mismatches 204; Indels 53; Gaps 11;
QY 52 TWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKIMMERKEHTNLMST 105
```

```
DB 13 SWERGVLGDQLVSDNELQEMSTQSGKYIDREIQNAVKGVEIKTIETNEERKTLISV 72
QY 106 LKKCREKEQEAALKLLNEVQEHLEEEERLCRESLADSWGECRSLNENCMRIYT-TCQPSW 164
DB 73 LEEAKNKEDALNETRDSETKLKAPEVCNETMMALWEECKPCLKQTKMFKYARVCRSGS 132
QY 165 SSVKNNKIERFFKIQYQFLFPFHEDNEKOLPISEKLEKDAQLTQMEDVFSQLTVDVNSLF 224
DB 133 GLVGRLQEEFLNQSSPFYEWINGDR-----IDSLLENDRQQSHVLDV-----MQDSF 179
QY 225 NRSENVFROMQQEFQDTQSFHSIFSDPLTEPYFFPAFSKEPMTKADLEOCWDIP----- 278
DB 180 NRATGI---MDELFDQDRFFTHKPDQTFYHSPFSY---FRPPPLHYAKSLRVNIMPISLYG 234
QY 279 --NFFQFCNFSVSIYESVETITMKLKAIEDLPKQD---KAPDHGGLISKMLPGQDRGL 333
DB 235 PLNFQDMQFPFFEMIHQAQOAMDVHLSPAYQTPNVEFITGGPD-----DRAV 282
QY 334 CGELDONLSRCFKFHEKCKOQAHLSDC-----PDVPAHLTELDEAIRLVNVSNOQYQI 389
DB 283 CREIRHNSGCLRMKDQCAKCEILSVDCSANNPSONQLRQELNDSLRLAELTKRYNEL 342
QY 390 LQWTRKHLEDATVLYEKMGRGQGWSELANQAPETETIIFNSIQVVPRIHEGNLSKQDETM 449
DB 343 LQYQWKMLNTSLLDQPNQFNWVSQLANLTQGPQYLYRVSTVTS--HTSEAPSRVT 401
QY 450 MTDLISLPSNFTLKIPLSEAESNFICYVYVAKALQHFKE 490
DB 402 EVVVKLFSDSPITITIPSEVRDNPKNFMETVAEKALQEYRK 442

RESULT 4
CLUS_CANFA
ID CLUS_CANFA STANDARD; PRT; 445 AA.
AC P25473;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Glycoprotein 80) (GP80).
GN CLU.
OS Canis familiaris (Dog).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
OX NCBI_TaxID=9615;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=91236776; PubMed=2033078;
RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pilarsky C.,
RA Appel D., Haase W., Mann K., Weller A., Koch-Brandt C.;
RT "Molecular cloning of gp 80, a glycoprotein complex secreted by
RT kidney cells in vitro and in vivo. A link to the reproductive system
RL J. Biol. Chem. 266:9924-9931(1991).
CC -!- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M55251; AAA30846.1; -
DR PIR; A40018; A40018.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
```

```
DR SMART: SM00035; CLa: 1.
DR SMART: SM00030; CLb: 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 445
FT CHAIN 23 445
FT BETA-CHAIN (P35).
FT ALPHA-CHAIN (P45).
FT INTERCHAIN (BY SIMILARITY).
FT DISULFID 102 309
FT DISULFID 113 301
FT DISULFID 116 298
FT DISULFID 121 291
FT DISULFID 129 281
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 277 277
FT CARBOHYD 287 287
FT CARBOHYD 350 350
FT CARBOHYD 370 370
SQ SEQUENCE ~445 AA; 51789 MW; 023A37266ABEF374 CRC64;

Query Match 15.6%; Score 410.5; DB 1; Length 445;
Best Local Similarity 23.9%; Pred. No. 5.1e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 30 MKPPLLVFVCLWLKDKSHCAPTKWKTATSENLSKSFSEVGEIDAD-----EEVKKALT 83
Db 1 MKKTLTLLVGLLL-----TWNGRVGLDQAVSDTELOEMSTEGSKYINKEIKALK 51
QY 84 GIKOMKIMRKEKEHTNLMSTLKKREKQEOALKLLNEVQHEEERLEECRESLADSWG 143
Db 52 GVGIKTLTLEOTNEERKSLNLEAKKEDALNDTKDSETKLKASQGVNDTFMMAWE 111
QY 144 ECRSCLENNCHRYT-TCQPSWSVKNKIERFPKIVQFLPPPHEDNEKDLPISEKLIK 202
Db 112 ECKPCLAKOTCMKFYARVCRSGSLVGHQLEEFNLQSSPFYFWMNGDR-----IDSLLEN 165
QY 203 DAQLTOMEDVFSQLTVDVNSLFNRSFNVFMQOEFDOTFQSHFIS----DITDTEPYFF 258
Db 166 DROOQTHADV-----MQDSFNRASSI-----MDELQDFRFTREPQDTHYSP--F 209
QY 259 PAFSKEPMTKADLEQCWDI-----PNFQLCFNFSVSIYSESVETITMLKAIB-- 307
Db 210 SLFQRPFPPFKFIARNIIPFRFQPLNFHDMFPQFPFDMHQ-A-QQADVNLHRIPIYH 268
QY 308 --DLPKODKAPDGHGLISKMLPGDRCGLGELDONLSRCFKFHEKCKQCAHLESDC--- 362
Db 269 PIEFPEED-----NRTVCKEIRHNSTGCLMKDKQCEKQOEILSVDCSSN 312
QY 363 -PDVPALHTELDAIRLVNVSNOQYGQILQWTRKHLDTAYLVKEMRGQFGWYSELANQA 421
Db 313 NPAQVQLRQELNSLQIAEFTKLYDELQSYQEKMENTSLKQLNEQSWSQLANLT 372
QY 422 PETEIIINSQVY-PRTHEGNISKQDQDTMTDLSILPSSNFTIKIPLEESAESSNFTGYV 480
Db 373 QSEDPFYLVQTVGTSQSDSNVPVGFVKV--VKLFSDSDPTVWIPAVSNPNPKFMEV 430
QY 481 VAKALQHFKE 490
Db 431 AERALKQEIYQ 440

RESULT 5
CLUS_MOUSE
ID CLUS_MOUSE STANDARD; PRT; 448 AA.
AC Q06890;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Clusterin)
```

```
DE (Apolipoprotein J) (Apo-J).
GN Clu OR MSGP-2 OR APOJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93356785; PubMed=8352774;
RA Lee K.-H., Ji Y.-M., Lim H.M., Lee S.-C., You K.-H.;
RT "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
RT from testis of mouse: implications of two different mRNAs of SGP-2.";
RL Biochem. Biophys. Res. Commun. 194:1175-1180(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93359508; PubMed=8354695;
RA French L.E., Chonn A., Ducrest D., Baumann B., Belin D., Wohlwend A.,
RA Kiss J.Z., Sappino A.P., Tschopp J., Schifferli J.A.;
RT "Murine clusterin: molecular cloning and mRNA localization of a gene
RT associated with epithelial differentiation processes during
RT embryogenesis.";
RL J. Cell Biol. 122:1119-1130(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
RA Strauch A.R.;
RL Submitted (XXX-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RC MEDLINE=94223204; PubMed=8169523;
RA Jordan-Starck T.C., Lund S.D., Witte D.P., Aronow B.J., Ley C.A.,
RA Stuart W.D., Swertfeger D.K., Clayton L.R., Sells S.F., Paigen B.;
RT "Mouse apolipoprotein J: characterization of a gene implicated in
RT atherosclerosis.";
RL J. Lipid Res. 35:194-210(1994).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A
CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
CC WITH PROGRAMMED CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC -1- TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
CC -1- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; D14077; BAA03162.1; -
DR EMBL; L08235; AAA37422.1; -
DR EMBL; L05670; AAA37284.1; -
DR EMBL; S70244; AAB30623.1; -
DR PIR; A40714; A40714.
DR PIR; JN0699; JN0699.
DR MGD; MGI:88423; Clu.
DR InterPro; IPR000753; Clusterin.
DR Pfam; PF01093; Clusterin; 1.
DR SMART; SM00035; CLa; 1.
DR SMART; SM00030; CLb; 1.
DR PROSITE; PS00492; CLUSTERIN_1; 1.
DR PROSITE; PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Sulfation; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 448 CLUSTERIN.
FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT)
```



```
Db 233 -----NFHDVFPF-----YDMIHQQAQAMDAHLQ-----RTPYHPTM-EFTENNDRTVC 277
Qy 335 GELQNLRSRCFKFEKCOKCOAHLSEDC-----PDVPAHLTDLDEAIRLVNVSNOYGOIL 390
Db 278 KEIRHNSTGGRMDKQCEKCEILEVDCSASNPQTLLRQOLNASLQAERFSLYDOLL 337
Qy 391 QMTRKHLEDATYLVKMRGQFGWSELANQAPETE-----IIFNSIQVVPRIHEGNISK 444
Db 338 QSYQOKMLNTSALLKQLNEQFTWVSQLANLTQSDQHYLVFTVNSHNSDPSIFSG----- 393
Qy 445 QDETMTDLSILPSNFTLKLPLESESSNFYGVVAKALQHK 489
Db 394 -----LTKVIVKLFNSFPITVTPQEVSSPNFMENVAEKALQYR 433

RESULT 7
CLUS_PIG STANDARD; PRT; 446 AA.
AC Q29549;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Clusterin precursor (Complement cytotoxicity inhibitor) (CLI).
GN CLI.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92184774; PubMed=1544909;
RA Diemer V., Hoyle M., Baglioni C., Millis A.J.;
RT "Expression of porcine complement cytotoxicity inhibitor mRNA in
RT cultured aortic smooth muscle cells. Changes during differentiation
RT in vitro."
RL J. Biol. Chem. 267:5257-5264(1992).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH (BY SIMILARITY).
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER (BY
CC SIMILARITY).
CC -1- TISSUE SPECIFICITY: HIGHEST LEVELS IN BRAIN AND LIVER; LOWER
CC LEVELS ARE DETECTED IN OTHER TISSUES, INCLUDING THE AORTA.
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: M84639; AAA31013.1; ..
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Cla; 1.
CC SMART: SM00030; Clb; 1.
CC PROSITE: PS00492; Clusterin_1; 1.
CC PROSITE: PS00493; Clusterin_2; 1.
KW Glycoprotein; Signal.
FT SIGNAL 1 28
FT CHAIN 29 446
FT CHAIN 29 227
FT CHAIN 29 446
FT CHAIN 228 446
FT CHAIN 102 312
FT DISULFID 113 304
FT DISULFID 116 301
FT DISULFID 121 294
FT DISULFID 129 284
FT CARBOHYD 86 85
FT CARBOHYD 103 103
FT CARBOHYD 103 103
BY SIMILARITY.
CLUSTERIN.
BETA-CHAIN (A CHAIN).
ALPHA-CHAIN (B CHAIN).
INTERCHAIN (BY SIMILARITY).
INTERCHAIN (BY SIMILARITY).
INTERCHAIN (BY SIMILARITY).
INTERCHAIN (BY SIMILARITY).
INTERCHAIN (BY SIMILARITY).
INTERCHAIN (BY SIMILARITY).
N-LINKED (GLCNAC...) (POTENTIAL).
N-LINKED (GLCNAC...) (POTENTIAL).
```

```
FT CARBOHYD 145 145 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 290 290 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 316 316 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 353 353 N-LINKED (GLCNAC...) (POTENTIAL).
FT CARBOHYD 373 373 N-LINKED (GLCNAC...) (POTENTIAL).
SQ SEQUENCE 446 AA; 51774 MW; BID5B434B668E3AA CRC64;

Query Match 15.2%; Score 400.5; DB 1; Length 446;
Best Local Similarity 24.4%; Pred. No. 2.1e-17;
Matches 118; Conservative 98; Mismatches 136; Indels 71; Gaps 16;

Qy 34 LLFIVICLLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDAEVKKALTKIK 86
Db 4 LLLVGLLL-----TWENGFWVLGDKAISDEKQEMSTEGSKYVKNKIKNALKEVK 54
Qy 87 QMKIMMERKEKHTNLMSTLKKCKREEKQEAALKLLNEQHELEERLERESLADSWGSCR 146
Db 55 QIKTLIEQSNEERKSLSSLEAKKKEDALNDTRDTETKLKSGQGLCNETMMALWESCK 114
Qy 147 SCLNNCMRIYT-TCQPSWSVKNKIERPKIYQFLPPFHEDNEKDLPISEKLIERDAQ 205
Db 115 PCLQOTCMKFYARVCRSGSLVGHQLEFLNQSSPFYFWINGDR-----IDSLMENDRQ 168
Qy 206 LTQMEDVFSQLTVDVNSLNFNSFNVRMQOEEDQTFQSHFIS-----DTDLTEPY----- 256
Db 169 QSHVMDI-----MEDSFNRASNI-----MDLQDFRFRNREPPTQTFSPFGSSHR 214
Qy 257 ---FFPAFSKEPMTKADLEQCWDIPNFQLFNFSVSIYESVETITKMLKAIEDLPKQD 313
Db 215 GSLFFNPKSRFARNIMPPLFTDL-NVHDMFQPF---FDMIHQQAQAMDAHLRHPVH- 268
Qy 314 KAPDHGGLISKMLP--QODRGLCGLDQNLRCRPFKEKCKQCAHLSEDC-----PDVPA 367
Db 269 -FPEAG-----VPENSNDRAVCKEIRHNSTGCLRMKDCQCEKREILSVDCSASNSQMQ 321
Qy 368 LHTELDEAIRLVNVSNOYGOILQMTKRKHLDEATYLVKMRGQFGWSELANQAPETII 427
Db 322 LRQELYSIQMAEAFSKLYDQLQSLQYQKMLNTSLLKQLNEQFSWYSQLANLTQNDRY 381
Qy 428 FNSIQVVPRIHEGNISKQDEMTDLSILPSSNFTLKLPLESESSNFYGVVAKALQ 487
Db 382 YLQVTTV-NHSGSDPSVPSPGLTKVVKLVFDSYPTLIIPQEVSS--DPKFMETVAEALQ 438
Qy 488 FKE 490
Db 439 YRQ 441

RESULT 8
CLUS_RAT STANDARD; PRT; 447 AA.
AC P05371;
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-FEB-1994 (Rel. 28, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid
DE glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM-
DE 2).
DE GN
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX NCBI_TaxID=10116;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.
RX MEDLINE=88000523; PubMed=3651384;
RA Collard M.W., Griswold M.D.;
RT "Biosynthesis and molecular cloning of sulfated glycoprotein 2
RT secreted by rat Sertoli cells."
RL Biochemistry 26:3297-3303(1987).
RN [2]
RP SEQUENCE FROM N.A.
```

RA Pineault J.M., Tenniswood M.;  
RL Submitted (XXX-1991) to the EMBL/Genbank/DBJ databases.  
RN [3]  
RP SEQUENCE FROM N.A.  
RX MEDLINE-89149740; PubMed-2920020;  
RA Bettuzzi S., Hilpakka R.A., Gilna P., Liao S.;  
RT "Identification of an androgen-regressed mRNA in rat ventral prostate  
RT as coding for sulphated glycoprotein 2 by cDNA cloning and sequence  
RT analysis.";  
RL Biochem. J. 257:293-296(1989).  
RN [4]  
RP SEQUENCE OF 22-51 AND 227-256.  
RX MEDLINE-88326333; PubMed-3415696;  
RA Cheng C.Y., Chen C.C., Feng Z., Marshall A., Bardin C.W.;  
RT "Rat clusterin isolated from primary Sertoli cell-enriched culture  
RT medium is sulfated glycoprotein-2 (SGP-2).";  
RL Biochem. Biophys. Res. Commun. 155:398-404(1988).  
RN [5]  
RP CHARACTERIZATION OF TRPM-2.  
RX MEDLINE-90134121; PubMed-2299741;  
RA Bandyk M.G., Sawczuk I.S., Olsson C.A., Katz A.E., Buttyan R.;  
RT "Characterization of the products of a gene expressed during  
RT androgen-programmed cell death and their potential use as a marker  
RT of urogenital injury.";  
RL J. Urol. 143:407-413(1990).  
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH.  
CC -!- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -!- DEVELOPMENTAL STAGE: EXPRESSED BY CELLS UNDERGOING PROGRAMMED  
CC DEATH AS A RESULT OF THE HORMONAL STIMULI OR A TRAUMATIC INSULT.  
CC -!- PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED  
CC CARBOHYDRATES.  
CC -!- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation  
CC at the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC  
CC  
CC EMBL; M16975; AAA1273.1; -  
DR EMBL; M64723; AAA42298.1; -  
DR EMBL; M64733; AAA42299.1; -  
DR EMBL; X13231; CAA31618.1; -  
DR PIR; A27205; A27205.  
DR PIR; A31575; A31575.  
DR PIR; B31575; B31575.  
DR PIR; A45890; A45890.  
DR PIR; S18491; S18491.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; Cla; 1.  
DR SMART; SM00030; Clb; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Sulfation; Glycoprotein; Spermatogenesis; Signal;  
FT SIGNAL 1 21  
FT CHAIN 22 447 CLUSTERIN.  
FT CHAIN 22 447 BETA-CHAIN (SMALL SUBUNIT).  
FT CHAIN 227 447 ALPHA-CHAIN (LARGE SUBUNIT).  
FT DISULFID 101 312 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 112 304 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 115 301 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 120 294 INTERCHAIN (BY SIMILARITY).  
FT DISULFID 128 284 INTERCHAIN (BY SIMILARITY).  
FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 290 290 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 327 327 N-LINKED (GLCNAC. . .) (PROBABLE).

FT CARBOHYD 353 353 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CARBOHYD 373 373 N-LINKED (GLCNAC. . .) (PROBABLE).  
FT CONFLICT 187 187 D -> H (IN REF. 1).  
SQ SEQUENCE 447 AA; 51375 MW; 9E2FA33E5E0C146E CRC64;  
  
Query Match 14.4%; Score 380; DB 1; Length 447;  
Best Local Similarity 23.5%; Pred. No. 3.6e-16;  
Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;  
  
QY 34 LLVIVCLLWLKDSHCAPTWKDKTAISENLKSGFVEID-----ADEVKKALTGIKQ 87  
Db 4 LLLVALL-----TWDNMGVLGEQEFSDNELQELSTQGSRYVNEIQNAVGVKH 54  
QY 88 MKIMMERKEHNTLMSTLTKCKREEKQEAALKLLNEVQEHLEEEERLCRSLADSWGECS 147  
Db 55 IKTLEKTNARERKSLNLEEAKKKKEGALDDTRDSEMKLKAFPEVCNETMMALWEECKP 114  
QY 148 CLENNCMRIYT-TCQPSWSSVKNKIERFFRKIYQFLFFPHEDNEKDLPISEKLIKDAQL 206  
Db 115 CLKHTCMKFYARVCRSGSLVGRQLQLEEFNQSSPFYFMNGDR-----IDSLLESDRQQ 168  
QY 207 TQ-----MEDVFSOLTVDNVNSLFNRSPNFRMQQEPDQTFQSHFISDTDLTEPYFPP--- 259  
Db 169 SQVLDAMQDSFTRASCIIDTLFQDRF--ETHEPQDI-----HHF-----SPMGFPK 214  
QY 260 ---AFSKEPMTKADLEQCWDIP-NFQPLCNFSVSIVSEYSETITKMLKAIEDLPKQDKA 315  
Db 215 PHFLYPKSLVRLSPLSHYGLSPHNPFPFMDIMHQAQAMDVLHSPALQFPDQDFL 274  
QY 316 PDHGLISLMLPGQ-DRGLCGELDQNLSCFPEHEKCKCOAHLSEDC-----PDVPAH 370  
Db 275 KE-----GEDDPTVCYKEIRHNSSTGLKMKGCERKCEILSVDCSTNNPAQANLRQ 324  
QY 371 ELDEAIRLVNSNOQYGOILQTRKHLEDTAYLVKMRQGFQGWVSELN--QAPETETIIF 428  
Db 325 ELNDSLOVAERLTQQYNELLSLQSKMLNTSSLLQELNDQFTVWSQLANTQGDQYLRV 384  
QY 429 NSIQVVRPIHEGNISKQDQETMTDLSILPSSNFTLKIPLEESAESSNFTGYGVAKALQHF 488  
Db 385 STVTT-----HSSDSEVPSPRVTEVVVKLFSDPITVVLPPEVSKDNPKFMDTVAEKALQ 440  
QY 489 K 489  
Db 441 R 441  
  
RESULT 9  
CLUS\_MESAU  
ID CLUS\_MESAU STANDARD; PRT; 191 AA.  
AC P14883;  
DT 01-APR-1990 (Rel. 14, Created)  
DT 01-APR-1990 (Rel. 14, Last sequence update)  
DT 01-OCT-1996 (Rel. 34, Last annotation update)  
DE Clusterin (Sulfated glycoprotein 2) (SGP-2) (Fragment).  
GN CLU.  
OS Mesocricetus auratus (Golden hamster).  
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Cricetinae;  
OC Mesocricetus.  
OX NCBI\_TaxID=10036;  
RN [1]  
RP SEQUENCE OF 24-123 FROM N.A.  
RX MEDLINE-89386721; PubMed-2780570;  
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;  
RT "Changes in brain gene expression shared by scrapie and Alzheimer  
RT disease.";  
RL Proc. Natl. Acad. Sci. U.S.A. 86:7260-7264(1989).  
RN [2]  
RP SEQUENCE OF 1-23 AND 124-191 FROM N.A.  
RA Duguid J.R., Bohmont C., Liu N., Tourtellotte W.W.;  
RL Submitted (AUG-1989) to the EMBL/Genbank/DBJ databases.  
CC -!- FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A VARIETY



```

RESULT 10
MYS2_SCHPO
ID MYS2_SCHPO STANDARD; PRRT; 1526 AA.
AC Q9USI6; P78969;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Myosin type II heavy chain 1.
GN MYO2 OR SPC6545.05C.
OS Schizosaccharomyces pombe (Fission yeast).
OC Eukaryota; Fungi; Ascomycota; Schizosaccharomycetes;
OC Schizosaccharomycetales; Schizosaccharomycetaceae;
OC Schizosaccharomycetes.
OX NCBI_TaxID=4896;
[1]
RN SEQUENCE FROM N.A., AND FUNCTION.
RP STRAIN=972;
RC MEDLINE=98075862; PubMed=9415380;
RX May K.M., Watts F.Z., Jones N., Hyams J.S.;
RT "type II myosin involved in cytokinesis in the fission yeast,
FT Schizosaccharomyces pombe.";
RL Cell Motil. Cytoskeleton 38:385-396(1997).
[2]
RN SEQUENCE FROM N.A.
RP STRAIN=972;
RC Wood V., Rajandream M.A., Barrell B.G., Rieger M.;
RA Submitted (MAR-1999), to the EMBL/GenBank/DBJ databases.
RL

```

```

CC  -!- FUNCTION: REQUIRED FOR CELL DIVISION. IT IS A COMPONENT OF THE
CC  CDC12 'SPOT', A STRUCTURE THOUGHT TO MARK THE SITE OF SEPTATION.
CC  MAY WORK IN CONJUNCTION WITH MYO3.
CC  -!- SUBUNIT: BINDS TO CDC4 AND RLC1.
CC  -!- SIMILARITY: CONTAINS 1 MYOSIN-LIKE GLOBULAR HEAD DOMAIN.
CC  -!- SIMILARITY: CONTAINS 1 IQ DOMAIN.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation on
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
CC  ENBL; U75357; AAC49908.1; -.
CC  EMBL; AL049498; CAB39901.1; -.
CC  HSP; P08799; INMD.
CC  InterPro; IPR000048; IQ.
CC  InterPro; IPR004009; Myosin_N.
CC  InterPro; IPR001609; myosin_head.
CC  Pfam; PF00612; IQ; 1.
CC  Pfam; PF00063; myosin_head; 1.
CC  Pfam; PF02736; myosin_N; 1.
CC  PRINTS; PR00193; MYOSINHEAVY.
CC  ProDom; PD000355; myosin_head; 1.
CC  SMART; SM00015; IQ; 1.
CC  SMART; SM00242; MYSC; 1.
CC  PROSITE; PS50096; IQ; 1.
CC  Myosin; Actin-binding; ATP-binding; Coiled coil; Calmodulin-binding;
CC  Alkylation.
CC  FT DOMAIN 1 757 MYOSIN HEAD-LIKE.
CC  FT DOMAIN 758 787 IQ.
CC  FT DOMAIN 875 1244 COILED COIL (POTENTIAL).
CC  FT NP_BIND 170 177 ATP (POTENTIAL).
CC  FT DOMAIN 634 656 ACTIN-BINDING (BY SIMILARITY).
CC  FT DOMAIN 734 748 ACTIN-BINDING (BY SIMILARITY).
CC  FT MOD_RES 674 674 ALKYLATION (BY SIMILARITY).
CC  FT CONFLICT 1337 1337 S -> R (IN REF. 1).
CC  SEQUENCE 1526 AA; 176430 MW; D71D51D6578192BA CRC64;
CC  -----
Query Match 5.7%; Score 151; DB 1; Length 1526;
Best Local Similarity 20.6%; Pred. No. 0.095;
Matches 98; Conservative 73; Mismatches 198; Indels 106; Gaps 17;

Qy 65 SFSEVGEIDADEYVKKALTGIGKIMMERKEKHTNLMSTLKKCREKQAKLLNVEQ 124
Db 895 SFSETKQ--QENLQRESASLKNNLELSLEKTSKVTETL-----LSQENELK 942
Qy 125 EHLEERLRCESLADSGEGRCSLENNCMRIYTCQPSWSVKNKIEREFKTIQFLFP 184
Db 943 EKLSLEEK-----DLLDTKGELESRENN-----ATVLESEAEF----- 976
Qy 185 FHEDNEKDLPISEKLIERDAQLTOMEDYFSQITVDVNS--LFNRSFNVFQMQEFD--- 239
Db 977 ---NEQCKSLQETIVTKDAELDKLTYSIDYKTEIQEMRLTNQKNME-KSIQEGSLSE 1031
Qy 240 -----QTQSHFISDITDEP-----YFPAPSKPEMT 267
Db 1032 SLKRVKRLERENSTLSDVSLKQOKELSVLKGVQELTINNLEKVNLEADVKQLPKL 1091
Qy 268 KADLEQCHWDIPNFOLFQCNFSVSYSEITKMLKAIEDLP-KQDKAPDHGGLISKML 326
Db 1092 KKELESNDKQOLYQATKNKEAKVKCECLNNFKTSKTKELNKEERQCNLSDSLKYI 1151
Qy 327 PGQDRGLCGELDQNL-----SRCFHFKEKQCQAHLSDCDPVPALPHTEIDELRVNVS 382
Db 1152 ELQ-----ETHENLLKVSLENKKYKEGLEQLDL-EGLDKVDTFNFEKSKHRDLTFN 1204
Qy 383 NO-----QYQILQMTKRKHLDPAYLVEKMRGQGWVSELANOAPETEIFNSIQVVP 435
Db 1205 HESLUROSASYKEKLSIASSENKDLNNKVSSTLTQKNVLESPKASKVPELE-----RKITT 1259

```



```

DR InterPro: IPR001609; myosin_head.
DR Pfam: PF00612; IQ; 2.
DR Pfam: PF00063; myosin_head; 1.
DR Pfam: PF02736; Myosin_N; 1.
DR Pfam: PF01576; myosin_tail; 1.
DR PRINTS: PR00193; MYOSINHEAVY.
DR ProDom: PD000355; myosin_head; 1.
DR SMART: SM00015; IQ; 1.
DR SMART: SM00242; MYSC; 1.
DR PROSITE: PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ.
FT DOMAIN 843 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.5%; Score 145.5; DB 1; Length 1938;
Best Local Similarity 20.7%; Pred. No. 0.27;
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 59 ISENKSFSEVGEIDAEDEV-KVLTGKIMMERKEHEHNLMTLKKCR-----EEK 113
DB 1150 ISELEAS--GATSAQIENKREAFKMRRLDEATLQHEATAATLRKKQADSVAEL 1207
QY 114 QEALKLLNEQVHELEERERLCRESLAD--SWGECRSLENNCMRIYTCQPSSSVKNKI 171
DB 1208 GEQIDNLRVKQLEKESEKWEIDMNASIKSNIERTCTVQEQSEIKAKD 1267
QY 172 ERFRFKIYQFLFPFHEDNEKDLPISEKLIIEKDAQLTQMEDVFSQLTVDVNSLFRNSPNVF 231
DB 1268 EQQTQLI-----HDLNMQKARLTQNGELSHRVEKESESLISQLTSKQALTOQLLELK 1320
QY 232 RQMQQEFD-QTFQSHFIS-----DTDLTPEYFFAFKPEMTKADLEOCW-----DIPNFF 281
DB 1321 ROMEETKAKNAHAHQAQSSRHQCDLLR-----EQYEEQEAQKAEQALSKANSEVAQWK 1376
QY 282 QLCNFSVSVIYESVETITKMLKAIEDL-PKQDKAPDHGLSK---MLPGQDRLGGLGEL 337
DB 1377 TKYETDAIQTEELEEAQKKLAQRLQAEKTEFANSKASLEKTKQRLQGEVEDLMDRL 1436
QY 338 DQNLRSRC-----FKFHEKQCKQAHLSDCPDVPAHHTLDELARLVNVS 383
DB 1437 ERSHTACATLDKORNFQKVLAEWKQKLDSEQAEEAAQKESRSLSTEL---FKMRNAYE 1493
QY 384 QOYQGI--LQWTKHLEDYALVEKMRQGFQWSELANOAPET-----EII 427
DB 1494 EVDQETLRENNKQEE-----ISDLTEQIAETGKNIQAEKTKKLVQEE 1540
QY 428 FNSIQVPRTHEGNISKQDETMM---TDLGILPSSNFTLKIPLSEASNESSNFIYGVVAKA 484
DB 1541 KSDQVALEEVGSLSEHSKILRVQLELSQVKSSELDKRVKIEKDETEQLKRNSQRAAEA 1600
QY 485 LQ 486
DB 1601 LQ 1602

RESULT 13
RA50_METJA STANDARD; PRT; 1005 AA.
AC Q58718;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)

```

```

DE DNA double-strand break repair rad50 ATPase.
GN RAD50 OR MJ1322.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;
OC Methanococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kierulff A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The
CC rad50/mrell complex possesses single-strand endonuclease activity
CC and ATP-dependent double-strand-specific exonuclease activity.
CC Rad50 provides an ATP-dependent control of mrell by unwinding
CC and/or repositioning DNA ends into the mrell active site (By
CC similarity).
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U67572; AAB99331.1; -
CC TIGR: MJ1322; -
DR InterPro: IPR003439; ABC_transporter.
DR InterPro: IPR001687; ATP_GTP_A.
DR InterPro: IPR003405; SMC_C.
DR InterPro: IPR003395; SMC_N.
DR Pfam: PF02483; SMC_N; 1.
DR Pfam: PF02463; SMC_C; 1.
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.
FT NP_BIND 32 39 ATP (BY SIMILARITY).
FT DOMAIN 158 849 COILED COIL (POTENTIAL).
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.4%; Score 141.5; DB 1; Length 1005;
Best Local Similarity 20.0%; Pred. No. 0.22;
Matches 81; Conservative 70; Mismatches 131; Indels 123; Gaps 16;

QY 54 KDTAISNL-----KSFSEVGEIDAEDEVKALTGKIMMERKEHEHNLMT 105
DB 419 QEKSKTEKNINDLETRINKLLEETKNIDI-ESIENSKKEIEKKVLENQKEIKLKK 477
QY 106 LKCKREEKQALKLLNEVQEHLEEEERLCRESLADSGECRSCL-----ENNCRIYTCQ 161
DB 478 LGEINSE---IKRLKKILDDELKEVE-----GKPLCKTPIDENKKMELINQHK 522
QY 162 PSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIIEKDAQLTQMEDVFSQLTVDVN 221
DB 523 TQLNNKYTELEENKIKREI-----EKDIEKLEKIDKEENKLTLYLEKQSQIE 574
QY 222 SLFNRSENVFQMQEQFDQFQSHFISDTLTPYFPFAPFSKEPMTKADLEQCWDIPNFF 281
DB 575 ELEKLNK-YEQLDLEINKKISNVVINGKPVDE-----ILE---DIKSQL 615
QY 282 QLCNFSVSVIYESVETITKMLKAIEDLPKQDKAPDHGLSKMLPGQDRLG 334

```

```
Db 616 NFKFNYNOYLSAVSLNSVDEGINRRIKEIENI-----VSGWNK----- 656
QY 335 GELDQNLSCRFKHEKQKQAHLSDCDPVPALHTELDE-----AIRLVNSNQ----- 384
Db 657 -----EKCREELNKLREDEIRNLRKDKLNELKKEKELIEIENRRSLKFD 702
QY 385 QYQOILQMTKRKULEDTAYLVKMRGQFGWSELANOAPETEIFN 429
Db 703 KYKEVLGLTEK-----LEELKNIKDGLLEIYN 729

RESULT 14
ID UTRN_HUMAN STANDARD; PRT; 3433 AA.
AC P46939;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 01-MAR-2002 (Rel. 41, Last annotation update)
DE Utrrophin (dystrophin-related protein 1) (DRP1) (DRP).
GN UTRN OR DMDL.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
SEQUENCE FROM N.A.
RX MEDLINE=93096045; PubMed=1461283;
RA Tinsley J.M., Blake D.J., Roche A., Fairbrother U., Riss J.,
RA Byth B.C., Knight A.E., Kendrick-Jones J., Suthers G.K., Love D.R.,
RA Edwards Y.H., Davies K.E.;
RT "Primary structure of dystrophin-related protein.";
RL Nature 360:591-593(1992).
RN [2]
X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS) OF 144-261.
RX MEDLINE=99141377; PubMed=9887274;
RA Keep N.H., Norwood F.L.M., Moores C.A., Winder S.J.,
RA Kendrick-Jones J.;
RT "The 2.0-A structure of the second calponin homology domain from the
actin-binding region of the dystrophin homologue utrophin.";
RL J. Mol. Biol. 285:1257-1264(1999).
CC -!- FUNCTION: MAY PLAY A ROLE IN ANCHORING THE CYTOSKELETON TO THE
PLASMA MEMBRANE (BY SIMILARITY TO DYSTROPHIN).
CC -!- SUBCELLULAR LOCATION: NEUROMUSCULAR JUNCTION.
CC -!- TISSUE SPECIFICITY: MUSCLE.
CC -!- SIMILARITY: STRONG, TO DYSTROPHIN.
CC -!- SIMILARITY: THE ACTIN-BINDING DOMAIN IS OF A TYPE FOUND IN MANY
ACTIN-BINDING PROTEINS (SUCH AS ACTININ, DYSTROPHIN, FIMBRIN,
ABP-120, ABP-180, OR BETA-FODRIN).
CC -!- SIMILARITY: CONTAINS 2 CALPONIN-HOMOLOGY (CH) DOMAINS.
CC -!- SIMILARITY: CONTAINS 1 WW DOMAIN.
CC -!- SIMILARITY: CONTAINS 20 SPECTRIN REPEATS.
CC -!- SIMILARITY: CONTAINS 1 Z2-TYPE ZINC FINGER.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb.ch/announce/
or send an email to license@isb-sib.ch).
CC -----
DR EMBL; X69086; CA448829.1; -
DR PIR; S28381; S28381.
DR PDB; 1BHD; 16-FEB-99.
DR MIM; 128240; -
DR InterPro; IPR001589; Actinin_act_bind.
DR InterPro; IPR001715; Calponin_hom.
DR InterPro; IPR002017; Spectrin.
DR InterPro; IPR003122; TarH.
DR InterPro; IPR001202; WW.
DR InterPro; IPR002349; WW_domain.
DR InterPro; IPR000433; ZnF_Z2.
```

```
DR Pfam; PF00307; CH; 2.
DR Pfam; PF00435; Spectrin; 19.
DR Pfam; PF00397; WW; 1.
DR Pfam; PF00569; Z2; 1.
DR PRINTS; PR00403; WWDOMAIN.
DR SMART; SM00033; CH; 2.
DR SMART; SM00150; SPEC; 18.
DR SMART; SM00319; TarH; 1.
DR SMART; SM00456; WW; 1.
DR SMART; SM00291; ZnF_Z2; 1.
DR PROSITE; PS00019; ACTININ_1; 1.
DR PROSITE; PS00020; ACTININ_2; 1.
DR PROSITE; PS00021; CH; 2.
DR PROSITE; PS01159; WW_DOMAIN_1; 1.
DR PROSITE; PS00020; WW_DOMAIN_2; 1.
DR PROSITE; PS01357; ZF_Z2_1; 2.
DR PROSITE; PS01357; ZF_Z2_2; 1.
KW Structural protein; Actin-binding; Calcium-binding; Cytoskeleton;
KW Repeat; 3D-structure; Zinc-finger.
FT DOMAIN 1 246 ACTIN-BINDING.
FT DOMAIN 31 135 CH 1.
FT DOMAIN 150 252 CH 2.
FT REPEAT 253 308 SPECTRIN 1.
FT REPEAT 309 417 SPECTRIN 2.
FT REPEAT 418 526 SPECTRIN 3.
FT REPEAT 541 637 SPECTRIN 4.
FT REPEAT 687 798 SPECTRIN 5.
FT REPEAT 803 902 SPECTRIN 6.
FT REPEAT 1016 1083 SPECTRIN 7.
FT REPEAT 1125 1230 SPECTRIN 8.
FT REPEAT 1248 1344 SPECTRIN 9.
FT REPEAT 1432 1541 SPECTRIN 10.
FT REPEAT 1544 1649 SPECTRIN 11.
FT REPEAT 1652 1753 SPECTRIN 12.
FT REPEAT 1910 1968 SPECTRIN 13.
FT REPEAT 1976 2081 SPECTRIN 14.
FT REPEAT 2258 2333 SPECTRIN 15.
FT REPEAT 2399 2440 SPECTRIN 16.
FT REPEAT 2443 2556 SPECTRIN 17.
FT REPEAT 2559 2636 SPECTRIN 18.
FT REPEAT 2658 2688 SPECTRIN 19.
FT REPEAT 2691 2797 SPECTRIN 20.
FT DOMAIN 2812 2845 WW.
FT ZN_FING 3064 3111 Z2-TYPE.
SQ SEQUENCE 3433 AA; 394488 MW; EAEBDB409F858E5B CRC64;

Query Match 5.4%; Score 141.5; DB 1; Length 3433;
Best Local Similarity 22.2%; Pred. No. 0.9; Mismatches 176; Indels 83; Gaps 18;
Matches 94; Conservative 71;

QY 45 KDSHCAPTWKDKTAISENLKSFSEVGEIDADEVKKALTIGIKQKIMMERKEKHTNLS 104
Db 1551 KEAASLSEWLSATATELVQKSTSEGLGLDLDTEISWAKNVKLDLE-----KKKADLNTIT 1605
QY 105 TLKKREKQKALKLLNVEQHEERLCRESLADSGECSCLENCMRYITTCQPSW 164
Db 1606 -----ESSAALQNLEGSEPI-LEERLC--VLNAGWSRVRTWTDWCNTL----- 1647
QY 165 SSVKNKIERFFRKI-----YQFLFPFHEDNEKDLPISEKLEK-----DAQLTQMED 211
Db 1648 MHQWOLEIFDGNVAHISTWLYQAEALDDEIEKKPTSKQEEIVKRLVSELDANL-QVEN 1706
QY 212 VFSQLTVDVNSLNSFNVMQOQEFDTQF--SHFISDTDLTPEYFFPAFSKEPMTKA 269
Db 1707 VRDQALILMNARGSSSRELVEPKLAELNRNFEKVSHIKSAKL-----LIAQEP---- 1755
QY 270 DLEQGWDPNFFQLFCNFSVIYESVETITKMLKAI-----BDLPKQKAPDHGLSKM 325
Db 1756 -LYQCLVTTETFTGTGVPF--SDLEKLENDIENMLKFEVKEHLESSEDEKMDSESAQIEV 1812
QY 326 LPQDGRGLCGELDQNLSCRFKHEKQKQAHLSDCDPVPALHTELDEAIRLVNSNQ 385
```

Db 1813 LQGEEMLHQPMDN-----KKEIKRLQLL-----LHTRYNK-IKAIPQQRK 1855  
 QY 386 YGQLQWTRKHLEDYAVLVKMRGQFGWSELANQAPETEIEFNSIQVVPRIHEGNSIKQ 445  
 Db 1856 MGQLASGIRSLPTDYLVE-----INKILLCHDDVLSLNVPELNTAIVE-DFSQ 1906  
 QY 446 DETM 449  
 Db 1907 EDSL 1910

RESULT 15  
 CENE\_HUMAN STANDARD; PRT; 2663 AA.  
 AC Q02224;  
 DT 01-JUL-1993 (Rel. 26, Created)  
 DT 01-JUL-1993 (Rel. 26, Last sequence update)  
 DT 30-MAY-2000 (Rel. 39, Last annotation update)  
 DE Centromeric protein E (CENP-E protein).  
 GN CENPE.  
 OS Homo sapiens (Human).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 OX NCBI\_TaxID=9606;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RX MEDLINE=93024922; PubMed=1406971;  
 RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;  
 RT "CENP-E is a putative kinetochore motor that accumulates just before  
 mitosis.";  
 RL Nature 359:536-539(1992).  
 RN [2]  
 RP CHARACTERIZATION.  
 RX MEDLINE=95196755; PubMed=7889940;  
 RA Thorer D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;  
 RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed  
 microtubule motor.";  
 RL EMBO J. 14:918-926(1995).  
 RN [3]  
 RP CHARACTERIZATION.  
 RX MEDLINE=98437347; PubMed=9763420;  
 RA Chan G.K.T., Schaar B.T., Yen T.J.;  
 RT "Characterization of the kinetochore binding domain of CENP-E reveals  
 interactions with the kinetochore proteins CENP-F and HUBB1.";  
 RL J. Cell Biol. 143:49-63(1998).  
 CC -1- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE  
 KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE  
 OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT  
 AND/OR SPINDLE ELONGATION.  
 CC -1- SUBUNIT: INTERACTS WITH CENP-F AND BUBR1 KINASE.  
 CC -1- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING  
 CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS  
 QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.  
 CC -1- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.  
 CC  
 CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 CC  
 DR EMBL: Z15005; CAA78727.1; -;  
 DR PIR: S28261; S28261.  
 DR HSPSP; P17119; 3KAR.  
 DR MIM; 117143; -;  
 DR InterPro: IPR001752; kinesin.  
 DR Pfam: PF00225; kinesin; 1.  
 DR PRINTS: PR00380; KINESINHEAVY.  
 DR SMART: SM00129; KISC; 1.  
 DR PROSITE: PS00411; KINESIN\_MOTOR\_DOMAIN1; 1.  
 DR PROSITE: PS00067; KINESIN\_MOTOR\_DOMAIN2; 1.

KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;  
 KW Cell cycle; Centromere.  
 FT DOMAIN 1 335 KINESIN-MOTOR.  
 FT COILED COIL (POTENTIAL).  
 FT DOMAIN 336 2471  
 FT GLOBULAR (POTENTIAL).  
 FT NP\_BIND 2472 2663  
 FT ATP (BY SIMILARITY).  
 SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8B8 CRC64;  
 Query Match 5.3%; Score 139.5; DB 1; Length 2663;  
 Best Local Similarity 18.1%; Pred No. 0.89;  
 Matches 100; Conservative 87; Mismatches 175; Indels 191; Gaps 23;  
 QY 59 ISENLSFSEVGSEIDAEVVKALTGK-----QMKIMMERKEKHTNLMSTLK----- 107  
 Db 1461 LKENIKEI-VAKHLETEELKVAHCHLKEQETINELRVNLSKETEISTIOKLEAIND 1519  
 QY 108 ----KCRE--EKQAL--KLLNEVQEHLEE-----PERLCRESLADSWGECRSCLNNCM 154  
 Db 1520 KLQNKIQETIYEKEEQNLNIQISEVQENVNELKQFKHRKADKALSQ-----IESKML 1572  
 QY 155 RIVTTCQPSWSSVK-----NKKIERFFPKIYQFL- 182  
 Db 1573 ELNRLQESQEEIQIMKEEMKRVQEQALQIERDQKENTKEIVAKMKESQEKYQFLK 1632  
 QY 183 -FPFHEDNEKDLPISEKLIKDAQLTQMEDVFSQLTVDVNSLFRSFNVFROMQOEFDO 241  
 Db 1633 MTAVNETQEKMCIEHLKEQFETQKLNLEIETE-NIRLTQILHENLEEMRSVTKERD-- 1689  
 QY 242 FQSHFISDLDLTPYFPFAPAFSKEPMYKADLEQCWDIPNFQPCFNFSVSIYESVETIYK 301  
 Db 1690 -----DLR-----SVEETLKVVERDQ-----LKENLRETIIR 1715  
 QY 302 MLKAIEDLPKODK-----APDHGGLISKMLPGDQGLGELDQNLNLSRC----- 344  
 Db 1716 -----DLEKQEEKIVHMLKHEQETIDKL-----RGIVSEKTNELSNMOKDLEHSNDA 1764  
 QY 345 -----FKFHEKCKQCAHLSEDCPDVPA-----HTELDEAIRL 378  
 Db 1765 LKAQDLKIQEELRIAHMLKEQEQETIDKLGRIVSEKTDKLSNMOKDLENSNAKLOEKIQE 1824  
 QY 379 VNVSNQOYGOILQMTKRKHELDYAVLVKMRGQFGWSELANQAPETEIEFNSIQV----- 433  
 Db 1825 LKANEHQ-----LITLKKDVNETQKVSSEM-----EQLKKQKDOOSLTLKLEIENLNL 1873  
 QY 434 VPRIHEG-----NISKODETMTDLSILPSSNFTLKIPLESSESSNFYGVV 481  
 Db 1874 AQELHENLEEMKSVMKERDNLRRVEETIKLERD-----QLKESLOETKARDLEIQOEL 1926  
 QY 482 AKALQHFKEHFKT 494  
 Db 1927 KTARMLSKHEHKT 1939

Search completed: July 2, 2002, 11:58:14  
 Job time: 367 sec

**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:48:17 ; Search time 33.04 Seconds  
(without alignments)  
1439.595 Million cell updates/sec

Title: US-09-722-544A-2MOD  
Perfect score: 2632  
Sequence: 1 MKIKAEKNEGSPSRWQLHW.....FIGYVAKALQHFKHEFTW 495  
Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues  
Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71:\*  
1: pir1:\*  
2: pir2:\*  
3: pir3:\*  
4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	434	16.5	449	1 A41386	clusterin precursor
2	418	15.9	451	2 I50131	clusterin - quail
3	410.5	15.6	445	2 A40018	clusterin precursor
4	408	15.5	449	2 S07714	T64 protein precursor
5	403	15.3	448	2 A40714	clusterin precursor
6	402	15.3	448	2 I56335	apolipoprotein J -
7	400.5	15.2	439	2 A35744	clusterin precursor
8	400.5	15.2	446	2 A42108	clusterin precursor
9	380	14.4	447	2 A27205	clusterin precursor
10	168	6.4	191	2 I48174	sulfated glycoprotein
11	164	6.2	1738	2 T14867	interaptin - slime
12	151	5.7	1526	2 T41522	myosin ii - fission
13	150	5.7	1156	2 B70356	chromosome assembly
14	149	5.7	1388	2 S70633	serine/threonine-s
15	148.5	5.6	1091	2 T34107	hypothetical prote
16	147	5.6	3225	2 I52300	giantin - human
17	147	5.6	3259	1 A56539	giantin - human
18	145.5	5.5	594	1 A44073	C1K1 protein - yea
19	141.5	5.4	1005	2 A44465	hypothetical prote
20	141.5	5.4	3433	1 S28381	utrophin - human
21	140.5	5.3	2166	2 G70163	hypothetical prote
22	140	5.3	1132	2 T00259	hypothetical prote
23	139.5	5.3	2663	1 S28261	centromere protein
24	138	5.2	1084	2 G71329	hypothetical prote
25	135	5.2	1955	2 T30934	myosin-like protei
26	137	5.2	1940	1 S04090	myosin heavy chain
27	136.5	5.2	1410	1 A57013	early endosome ant
28	136	5.2	1388	2 S74245	serine/threonine-s
29	135.5	5.1	1679	2 S48385	hypothetical prote

30	135.5	5.1	1956	2 T16416	hypothetical prote
31	135	5.1	1060	1 A40264	kinesin-related pr
32	135	5.1	1133	2 T22976	hypothetical prote
33	134.5	5.1	1300	2 I53799	CG1 protein - huma
34	134	5.1	1538	2 T28095	cardiac muscle fac
35	133.5	5.1	1313	2 A48467	myosin heavy chain
36	133.5	5.1	1957	2 T38077	hypothetical coile
37	132.5	5.0	852	2 D72230	conserved hypotet
38	132	5.0	1033	2 T42701	hypothetical prote
39	132	5.0	1992	2 A47297	myosin heavy chain
40	131.5	5.0	1127	2 T28317	ORF MSV156 hypot
41	131.5	5.0	1979	1 S03166	myosin heavy chain
42	131	5.0	1024	2 T34517	kinesin-related pr
43	131	5.0	1413	-2 T26467	hypothetical prote
44	130.5	5.0	971	2 D70128	conserved hypotet
45	130.5	5.0	1940	1 A24922	myosin heavy chain

ALIGNMENTS

RESULT 1

A41386

clusterin precursor [validated] - human

N: Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein

N: Contains: clusterin alpha chain; clusterin beta chain

C: Species: Homo sapiens (man)

C: Date: 03-Apr-1992 #sequence\_revision 17-Nov-1995 #text\_change 08-Dec-2000

C: Accession: S43646; S04662; A41386; A35833; S34036; A33177; A37816; B37816; PL0136; R: Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M. Eur. J. Biochem. 221, 917-925, 1994

A: Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with A: Reference number: S43646; MUID: 94237156

A: Accession: S43646

A: Molecule type: DNA

A: Residues: 1-449 <NON>

A: Cross-references: GB:M64722; NID:G339972; PIDN:AAB06508.1; PID:G339973 R: Kirszbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.; EMBO J. 8, 711-718, 1989

A: Title: Molecular cloning and characterization of the novel, human complement-associ A: Reference number: S04662; MUID: 89251601

A: Accession: S04662

A: Molecule type: mRNA

A: Residues: 1-449 <KIR>

A: Cross-references: EMBL:X14723; NID:G30250; PIDN:CAA32847.1; PID:G30251

A: Note: parts of this sequence, including the amino end of the mature protein, were c R: Jenne, D.E.; Tschopp, J. Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989

A: Title: Molecular structure and functional characterization of a human complement cy tis fluid.

A: Reference number: A41386; MUID: 89386692

A: Accession: A41386

A: Molecule type: mRNA

A: Residues: 2-449 <JEN>

A: Cross-references: GB:M25915; NID:G180619; PIDN:AAA35692.1; PID:G180620 R: de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gill, C.M.; Robbins, J. Biochemistry 29, 5380-5389, 1990

A: Title: Apolipoprotein J: structure and tissue distribution.

A: Reference number: A35833; MUID: 90344779

A: Accession: A35833

A: Molecule type: mRNA

A: Residues: 34-449 <DES>

A: Cross-references: GB:J02908; NID:G178854; PIDN:AAA51765.1; PID:G178855 R: Chiso, J.; Matsubara, E.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T. Biochem. J. 293, 27-30, 1993

A: Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's A: Reference number: S34056; MUID: 93319521

A: Accession: S34056

A: Molecule type: protein

A: Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI> R: James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse Arterioscler. Thromb. 11, 645-652, 1991

<p>A:Title: Characterization of a human high density lipoprotein-associated protein, NAL/NA</p> <p>A:Reference number: A53177; MUID:91230083</p> <p>A:Accession: A53177</p> <p>A:Molecule type: protein</p> <p>A:Residues: 229-242;303-304,'M',306-312,'X',314-317;397-403 &lt;JAM&gt;</p> <p>R:de Silva, H.V.; Stuart, W.D.; Park, Y.B.; Mao, S.J.T.; Gal, C.M.; Wetterau, J.R.; Busc</p> <p>J. Biol. Chem. 265, 14292-14297, 1990</p> <p>A:Title: Purification and characterization of apolipoprotein J.</p> <p>A:Reference number: A37816; MUID:90354412</p> <p>A:Accession: A37816</p> <p>A:Molecule type: protein</p> <p>A:Residues: 23-46,'H',48-51,'Q' &lt;DE&gt;</p> <p>A&gt;Note: amino end of the alpha chain</p> <p>A:Accession: B37816</p> <p>A:Molecule type: protein</p> <p>A:Residues: 228-257 &lt;DE2&gt;</p> <p>A&gt;Note: amino end of the beta chain</p> <p>R:Choi, N.H.; Mazda, T.; Tomita, M.</p> <p>Mol. Immunol. 26, 835-840, 1989</p> <p>A:Title: A serum protein SP40.40 modulates the formation of membrane attack complex of C</p> <p>A:Reference number: PL0135; MUID:90097955</p> <p>A:Accession: PL0135</p> <p>A:Molecule type: protein</p> <p>A:Residues: 23-37 &lt;CHO&gt;</p> <p>A&gt;Note: this fragment was isolated from the membrane attack complex SC5b-9</p> <p>A:Accession: PL0135</p> <p>A:Molecule type: protein</p> <p>A:Residues: 228-242 &lt;CH2&gt;</p> <p>A&gt;Note: this fragment was isolated from the membrane attack complex SC5b-9</p> <p>R:Hochstrasser, A.C.; James, R.W.; Martin, B.M.; Harrington, M.; Hochstrasser, D.; Pomet</p> <p>Appl. Theor. Electrophor. 1, 73-76, 1988</p> <p>A:Title: HDL particle associated proteins in plasma and cerebrospinal fluid: Identification</p> <p>A:Reference number: S07433; MUID:91265608</p> <p>A:Accession: S07433</p> <p>A:Molecule type: protein</p> <p>A:Residues: 229-240 &lt;HOC&gt;</p> <p>A:Accession: S07433</p> <p>A:Molecule type: protein</p> <p>A:Residues: 24-27,'S',29-33 &lt;HO2&gt;</p> <p>R:Kriszbaum, L.; Bozas, S.E.; Walker, I.D.</p> <p>FEBS Lett. 297, 70-76, 1992</p> <p>A:Title: SP-40.40, a protein involved in the control of the complement pathway, possesses</p> <p>A:Reference number: A56293; MUID:92201397</p> <p>A:Accession: A56293</p> <p>A:Molecule type: protein</p> <p>A:Residues: 229-240 &lt;HOC&gt;</p> <p>A:Accession: S07433</p> <p>A:Molecule type: protein</p> <p>A:Residues: 228-246 &lt;KUN&gt;</p> <p>A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma</p> <p>A:Accession: F34223</p> <p>A:Molecule type: protein</p> <p>A:Residues: 23-34,'X',36-37 &lt;KU2&gt;</p> <p>A:Experimental source: apolipoprotein A-I-containing lipoproteins, plasma</p> <p>R:Danik, M.; Chabot, J.G.; Mercier, C.; Benabid, A.L.; Chauvin, C.; Quirion, R.; Suh, M.</p> <p>Proc. Natl. Acad. Sci. U.S.A. 88, 8577-8581, 1991</p> <p>A:Title: Human gliomas and epileptic foci express high levels of a mRNA related to rat t</p> <p>A:Reference number: I59206; MUID:92020896</p> <p>A:Accession: I59206</p> <p>A:Molecule type: mRNA</p> <p>A&gt;Status: preliminary; translated from GB/EMBL/DBJ</p> <p>A:Molecule type: mRNA</p> <p>A:Residues: 61-449 &lt;RES&gt;</p> <p>A:Cross-references: GB:W74816; NID:g338056; PIDN:AAA60321.1; PID:g338057</p> <p>R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.</p> <p>Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989</p> <p>A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.</p> <p>A:Reference number: I48174; MUID:89386721</p> <p>A:Accession: I63132</p> <p>A&gt;Status: preliminary; translated from GB/EMBL/DBJ</p> <p>A:Molecule type: mRNA</p> <p>A:Residues: 103-168 &lt;RE2&gt;</p>	<p>A:Cross-references: GB:M26639; NID:g338070; PIDN:AAA36609.1; PID:g553644</p> <p>C:Comment: This protein has been implicated in complement cascade inhibition, membran</p> <p>11 as in normal brain, in tissues affected by neurodegenerative disease processes, an</p> <p>C:Comment: This protein may assist in preventing the formation of Alzheimer's disease</p> <p>C:Genetics:</p> <p>A:Gene: GDB:CLU; CLI</p> <p>A:Cross-references: GDB:125226; OMIM:185430</p> <p>A:Map position: 8p21-8p21</p> <p>A&gt;Note: appears to be a single-copy gene; alternative exon usage in 5'-untranslated r</p> <p>C:Superfamily: clusterin</p> <p>C:Keywords: apoptosis; complement inhibitor; extracellular protein; glycoprotein; HDL</p> <p>F:1-22/Domain: signal sequence #status predicted &lt;SIG&gt;</p> <p>F:23-227/Domain: clusterin beta chain #status experimental &lt;BCH&gt;</p> <p>F:228-449/Domain: clusterin alpha chain #status experimental &lt;NAT&gt;</p> <p>F:228-449/Domain: clusterin alpha chain #status experimental &lt;ACH&gt;</p> <p>F:86-103,145,291,354,374/Binding site: carbohydrate (Asn) (covalent) #status experime</p> <p>F:102-313,121-295,129-285/Disulfide bonds: #status experimental</p> <p>F:113-305,116-302/Disulfide bonds: (or 113-302, 116-305) #status experimental</p> <p>F:317/Binding site: carbohydrate (Asn) (covalent) #status absent</p> <p>Query Match 16.5%; Score 434; DB 1; Length 449;</p> <p>Best Local Similarity 25.8%; Pred No. 7.3e-20;</p> <p>Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;</p> <p>QY 30 MKPPLLVTIVCLLWKDSHCAPTWK-----DKTAISENLKSFSEVEIDADAEVKKALT 83</p> <p>Db 1 MMKTLTLFVGLL-----TWESGQVLDGQTVSDNELQMSNQSGSVYKNEIQNAVN 51</p> <p>QY 84 GTKQKIMMERKEKEHTNLMSTLKCRBQKQALKLLNEVQHLBEERLCRESLADSG 143</p> <p>Db 52 GVKQIKTLIEKTNEBKRTLLSLEAKKKKDALNETRESETKLKEPVCNETMMALWE 111</p> <p>QY 144 ECRSCLENNCRITYT--TCOPSSSVKNTIERFRKIYQFLFPFHEDNEKDLPISEKLEK 202</p> <p>Db 112 ECKPCLKQTCMKFYARVCRSGSLVGRQLEELNLSPPYFWMGDR-----ISLLEN 165</p> <p>QY 203 DAQLTQMEDV-----FSQLTVDVNSLFRNSFNVRMQQEDFQDTFSHFIS--TDTJTEPYF 257</p> <p>Db 166 DRQQTMLDVMQDHFSSRASSIIDELFQDRF-----FTREPQDTY--HYLPFSLPHRRPHF 218</p> <p>QY 258 RPAFSKEPMTKADLE--QCWDIDNFTQLCNFSVSIVSEYITKMLKAIEDLPQDKAP 316</p> <p>Db 219 F--FPKSRIVRSLMPSPSPYLFNFHAMFQPFLEIHEA-----QOAMDIFHFSAPF 267</p> <p>QY 317 DHGGLSKMLPQDRLGCLGELQNLRSRCFKFHEKQCKQAHLSDC----PDVPALHTBL 372</p> <p>Db 268 QHPPTFEIREGDDRTVCREIRHNSGCLRMKQDKCKREILSVDCSTNNPSQAKLRRL 327</p> <p>QY 373 DEAIRLVNSQOYQIILQMTKRKHELDYALVEKMRGQGVVSELANQAPETEIFNSIQ 432</p> <p>Db 328 DESLQVAERLTKRYNELLSKYQMKMLNTSSLLEQLNEQFNWYSRLANTQGEDQYLYRV 387</p> <p>QY 433 VVPRIHGKISKODETMTDLSILPNSFTLKLPLEESAESSNFYGVVAKALQHF--KEH 491</p> <p>Db 388 TVAS--HTSDSDVPSGVTEVVVVKLFSDPTITVTPVEVSRKPKFMETVAERKALQETRRKH 446</p> <p>RESULT 2</p> <p>150131</p> <p>clusterin - quail</p> <p>C:Species: Coturnix coturnix (quail)</p> <p>C:Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 13-Aug-1999</p> <p>C:Accession: 150131</p> <p>R:Michel, D.; Chatelain, G.; Herault, Y.; Brun, G.</p> <p>Eur. J. Biochem. 229, 215-223, 1995</p> <p>A:Title: The expression of the avian clusterin gene can be driven by two alternative</p> <p>A:Reference number: 150131; MUID:95262670</p> <p>A:Accession: 150131</p> <p>A&gt;Status: preliminary; translated from GB/EMBL/DBJ</p> <p>A:Molecule type: DNA</p> <p>A:Residues: 1-451 &lt;MIC&gt;</p> <p>A:Cross-references: EMBL:X80760; NID:g520629; PIDN:CAA56733.1; PID:g520630</p>
---	--



-----EEVKNALI 83

QV 237 EEDOTEOSHETSOTNITPEV--EEDAFESKEDMTKADIEOCWDION-----EEYES 28E

Qv 237 F

237 EFDDTFOSHFISDTT

QY 237 EFDO7E0SHF7SD7DI.TEPV--EPPAESEK

0v 237 EEDOTEOSHETSDTNT.TEPY--EEDAFESKERMTKADIECQNDT

QV 237 EEDOTEOSHETSDDITTEPV--EEDAESEKPEBMTKADIEOCWQDIPN-----EEC

OV 237 EEDOTEOSHETSOTDITERY--EEDAESEKEDMTKADIEOQNDIDN-----EEOYEC 20E

```
Db 203 -----TPPFGGFEAF-VPPVQRVHL-----VPRRLSRELHPFHQPM 240
QY 286 NFSVSIYESVETITKMLKAIEDLPKQKADPH--GGLSKMLP-GQDRGLCGELDONLS 342
Db 241 HGFHRLFOPLFEMTQHMLDG-----GHGAWHEPLGGFATESRNFSTDRMVCREIRNSA 294
QY 343 RCFKFEHKCKQCAHLSEDC----PDVPALHTELDEARLVNVSNOQYGOIQLQMKRKHLE 398
Db 295 GCLMRDECEKREILAVDCSDTPVQSOLRQFEDALRAERFTRRYDDLLSAFQAEML 354
QY 399 DTAIVLEKMRGQFGWVSELANQAP-----ETEILFNISQVVPRIHEGNISKQDETMWD 452
Db 355 NTFSSLDQLNRFGWVSRLCNLTQNGDGLFTVTFESK----TPNLEDP--SAPADTQVT- 408
QY 453 LSLPSNFTLKIPLEESAESSNFYGVVAKALQHF 488
Db 409 VQLFDSLEPLSLVPGDISWDDPRFMEIVPEQALQHY 444

RESULT 5
A40714
Clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
C:Accession: A40714; JN0699
R:French, L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <FRE>
A:Cross-references: GB:L08253; NID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LEE>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-226,227-448/Product: clusterin #status predicted <MAT>
F:227-448/Domain: clusterin alpha chain #status predicted <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) #status predicted

Query Match 15.3%; Score 403; DB 2; Length 448;
Best Local Similarity 23.1%; Pred. No. 6.3e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

QY 34 LLVIFVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGIQK 87
Db 3 ILLLCVALLI-----WDNGVVLGQEVSDNELQELSTQGSRYINKEIQNAVQGVKH 54
QY 88 MKTMRKEKEHTNLMSTLKKREEKQKALKLLNEVQHELEERLCRESLADSGECRS 147
Db 55 IKTLEKTNAERKSLNSLEEAKKKEDALEDTROSEMKLKAFPEVCNETMMALWEECKP 114
QY 148 CLENNCMRIYV--TCQPSWSVKNKIERFRFKIYQFLFPFHEDNEKDLPISEKLIKDAOL 206
Db 115 CLKHTCKFYARVCRSGSLVGQOLEFLNQSSPFYFWMNGDR-----IDSLESRQ 168
QY 207 TQ-----MEDVFSQLTVDVNSLNRSNFVRMQQEQFDQTFQSHFISDTLTPEYFP 259
Db 169 SQVLDAQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGPPH 214
```

```
QY 260 ---AFSEKPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVETITKMLKAIEDLPKQDKA 315
Db 215 PHFLYPKRLSLKSPSHYGPSPFNMFQPFEMIHQAQAMQVQLHSPAFQFPDQV-- 272
QY 316 PDHGLISKMLPGQ--DRGLCGELDONLSRCFKFEHKCKQCAHLSEDC----PDVPALHT 370
Db 273 -----FLRGEEDRTVCKEIRNRNCGCLAMKQCEKQCEILSVDCSTNNPAQANLRQ 324
QY 371 ELDEAIRLVNVSNOQYGOIQLQMKRKHLEDTAYLVKMRGQFGWVSELANQAPETEIFNS 430
Db 325 ELNDSLQVAERLTEQYKELLQSFQSKMLNTSSLLEQLNDQFNWVSQLANLTQGEDKYLYR 384
QY 431 IQVVPRIHEGNISKQDETMMDLSILPSSNFTLKIPLEESAESSNFYGVVAKALQHF 489
Db 385 VSTV--TTHSSDSEVPVSRVTEVVVVKLFDSDPITVVLPEEYVKDNPKFMDTVAEKALQ 442

RESULT 6
156335
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
C:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Aronow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; NID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
A:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.3%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 7.3e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

QY 30 MKPPLLVTIVCLLWLKDSHCAPTWKDKTAISEN-LKSFSEVGEIDADEVKALTGIK 88
Db 1 MKILLVCVALLI--DNGW---LGEQVSDNEIQELSTQGSRYINKEIQNAVQGVKH 55
QY 89 KTMERKEKEHTNLMSTLKKREEKQKALKLLNEVQHELEERLCRESLADSGECRS 148
Db 56 KTLIEKTNAERKSLNSLEEAKKKEDALEDTROSEMKLKAFPEVCNETMMALWEECKP 115
QY 149 LENNCMRIYV--TCQPSWSVKNKIERFRFKIYQFLFPFHEDNEKDLPISEKLIKDAOLT 207
Db 116 LKHTCKMFYARVCRSGSLVGQOLEFLNQSSPFYFWMNGDR-----IDSLESRQ 169
QY 208 Q-----MEDVFSQLTVDVNSLNRSNFVRMQQEQFDQTFQSHFISDTLTPEYFP 259
Db 170 QVLDAQDSFARASGIIDTLFQDRF--FAR-----ELHDPHYFS-----PIGPPH 215
QY 260 ---AFSEKPMTKADLEQC-WDIPNFFQLFCNFSVSIYESVETITKMLKAIEDLPKQDKAP 316
Db 216 HFLYPKSLRVLSLSPSHYGPSPFNMFQPFEMIHQAQAMQVQLHSPAFQFPDQV-- 272
QY 317 DHGGLISKMLPGQ--DRGLCGELDONLSRCFKFEHKCKQCAHLSEDC----PDVPALHT 371
Db 273 -----FLUREGDDRTVCKEIRNRNCGCLAMKQCEKQCEILSVDCSTNNPAQANLR 325
QY 372 LDEAIRLVNVSNOQYGOIQLQMKRKHLEDTAYLVKMRGQFGWVSELANQAPETEIFNS 431
Db 326 LNDLSQVAERLTEQYKELLQSFQSKMLNTSSLLEQLNDQFNWVSQLANLTQGEDKYLYR 385
QY 432 QVVVPRIHEGNISKQDETMMDLSILPSSNFTLKIPLEESAESSNFYGVVAKALQHF 489
Db 386 STV--TTHSSDSEVPVSRVTEVVVVKLFDSDPITVVLPEEYVKDNPKFMDTVAEKALQ 442
```

## RESULT 7

A35744  
clusterin precursor - bovine  
N:Alternate names: complement cytolysis inhibitor; glycoprotein III  
N:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 13-Aug-1999  
C:Accession: A35744  
R:Palmer, D.J.; Christie, D.L.  
J. Biol. Chem. 265, 6617-6623, 1990  
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin granules  
A:Reference number: A35744; MUID:90216681  
A:Accession: A35744  
A:Molecule type: mRNA  
A:Residues: 1-439 <PAL>  
A:Cross-references: GB:J05391; NID:g163114; PID:AAA30554.1; PID:g163115  
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were determined from complementary DNA  
C:Superfamily: clusterin  
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-221/Domain: clusterin beta chain #status experimental <BCH>  
F:20-221,222-439/Product: clusterin #status experimental <MAT>  
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>  
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 15.2%; Score 400.5; DB 2; Length 439;  
Best Local Similarity 24.5%; Pred. No. 8.8e-18;  
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 52 TWKDKTAISE-NLKSFEVEGIDEADDEVKALGTGKIMMERKEKEHTNLMTSLKCCR 110  
Db 13 SWESGWAISKEIQEIMSTEGSKYVKNKEIKALKEVQIKQIEQNEERKLLLSLEAK 72  
QY 111 EKEQALKLLNEVOEHLERLCRESLADSGECRSCLENNCMRIYT-TCQPSWSSVKN 169  
Db 73 KKEDALNDRDSNKLKASGGVCNETWTALWEECKPCLKQTCMKFYARVCRSGSLGVH 132  
QY 170 KIERFRKIYQFLPPFHEDNEKOLPISEKLEKDALQTOEMEDVFSQLTVDVNSLFRSEN 229  
Db 133 QLEEFLLNQSPFYFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRASS 179  
QY 230 VFRMQQEFDTQSHFI-----SDTDLTEPY-----FF--PAPSKPEMTKADLEQC 274  
Db 180 I-----MDLEFQDRFLRRPDQTYSPFSFGRSLFNPKSRFARNVMPFLLEPF 232  
QY 275 WDPNPFQFCNFSVSIYESVSETITKMLKAIEDLPKODKAPDHGGGLISKMLPQDRLC 334  
Db 233 ---NPHDVFPF---YDHIHQAAQQAQDAHLQ-----RTPHFPTM-EFTENNDRVC 277  
QY 335 GELDONLRCFKFHEKCKQKCOAHLSEDC-----PDVPALHTDELDAIRLVNVSNQYQQL 390  
Db 278 KEIRHNSTGCLRMKQCEKQCEILEVDSCASNPTQLLRQOLNASLQAEKFSRLYDQL 337  
QY 391 QMTKRLHEDPAILVEKMRGQFVWSELANOAPETE-----IIFNSIQVVPRIHEGNISK 444  
Db 338 QSQQKMLNATSALEKQNEQFTWVSQLANTQSDQHQYLVQFTVNSHNSDPSPSG---- 393  
QY 445 QDEMTMTDLSILPSSNFTLKIPLEESAESSNFIYGVVAKALQHF 489  
Db 394 ----LTKVIVKLFNSFPITVTPQEVSSPNFMENVAEKALQOYR 433

## RESULT 8

A42108  
clusterin precursor - pig  
N:Alternate names: complement cytolysis inhibitor; CP40 protein  
C:Species: Sus scrofa domestica (domestic pig)  
C:Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C:Accession: A42108; JC5535; PC4475  
R:Dlemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.  
J. Biol. Chem. 267, 5257-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic endothelial cells  
A:Reference number: A42108; MUID:92184774  
A:Accession: A42108  
A:Molecule type: mRNA  
A:Residues: 1-446 <DIE>  
A:Cross-references: GB:M84639; NID:g164408; PID:AAA31013.1; PID:g164409  
A:Experimental source: aortic smooth muscle cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)  
Biochem. Biophys. Res. Commun. 234, 712-718, 1997  
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.  
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary gland  
A:Reference number: JC5535; MUID:97318844  
A:Accession: JC5535  
A:Molecule type: DNA  
A:Residues: 1-446 <OGA>  
A:Accession: PC4475  
A:Molecule type: protein  
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>  
A:Experimental source: pituitary gland  
C:Superfamily: clusterin  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-227/Domain: clusterin beta chain #status predicted <BCH>  
F:23-227,228-446/Product: clusterin #status predicted <MAT>  
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

Query Match 15.2%; Score 400.5; DB 2; Length 446;  
Best Local Similarity 24.4%; Pred. No. 9e-18;  
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 34 LLVFIIVCLWLKSDKCAPTWK-----DKATJSENLKSFSEVGEIDAEDEVKALTGK 86  
Db 4 LLLLVGLLL-----TWENGFVWLGDKAISDKELQEMSTEGSKYVKNKEIKALKEV 54  
QY 87 QMKIMMERKEHTNLMTSLKCRREEKQEAALKLLNEVOEHLERLCRESLADSGECR 146  
Db 55 QIKTLIEQSENEERKSLLSLEAKKKEDALNDRDTETKLGKSGQGLCNETMMALWEECK 114  
QY 147 SCLENNCMRIYT-TCQPSWSSVKNKIERFRKIYQFLPPFHEDNEKDLPISEKLEKDAQ 205  
Db 115 PCUKQTCMKFYARVCRSGSLGVHGLEEFLLNQSPFYFWINGDR-----IDSLMENDRQ 168  
QY 206 LTQMEDVFSQLTVDVNSLFRSENFRMQQEFDTQSHFIS-----DSDLTEPY----- 256  
Db 169 QSHVMDI-----MEDSFNRASNI-----MDLEFQDRFNRFPFDTQFSPFGSSHR 214  
QY 257 ---FFPAFKPEMTKADLEQCWDIPNFFQFCNFSVSIYESVSETITKMLKAIEDLPKOD 313  
Db 215 GSLFFNPKSRFARNIMPFLFTDL-NYHDMFQPF-----FDMIHQAQAQDAHLHRIYH- 268  
QY 314 KAPDHGGLISKMLP--GODRGLCGELDONLRCFKFHEKCKQKCOAHLSEDC-----PDVPA 367  
Db 269 -FPEAG-----VPENSDNRVCKEIRHNSTGCLRMKQCEKREILSDSCASNSSQMQ 321  
QY 368 LHTELDEAIRLVNVSNQYQQLQMTKRLHEDPAILVEKMRGQFVWSELANOAPETEII 427  
Db 322 LRQELYTSIQMAEKFSKLYDQLQSYQKMLNTSSLKQLNEQFSWVSQLANTQNDRY 381  
QY 428 ENSIQVVPRIHEGNISKQDETMMTDLNILPSSNFTLKIPLEESAESSNFIYGVVAKALQH 487  
Db 382 YLQVTTV-NSHGSDPSVPSGLTKVVVVKLFDSYPITLIIPOEVS--DPKFMETVAEALQO 438  
QY 488 FKE 490  
Db 439 YRQ 441

## RESULT 9

A27205  
clusterin precursor - rat  
N:Alternate names: SGP-2; sulfated glycoprotein 2; TRPM-2  
C:Species: Rattus norvegicus (Norway rat)  
C:Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 13-Aug-1999

C:Accession: A45890; S18491; A45415; B31575; A31575; A27205  
R:Collard, M.W.; Griswold, M.D.  
Biochemistry 26, 3297-3303, 1987  
A:Title: Biosynthesis and molecular cloning of sulfated glycoprotein 2 secreted by rat S  
A:Reference number: A45890; MUID:88000523  
A:Accession: A45890  
A:Molecule type: mRNA  
A:Residues: 1-447 <CO2>  
A:Cross-references: GB:M16975; NID:g204472; PIDN:AAA41273.1; PID:g204473  
R:Bettuzzi, S.; Hlipakka, R.A.; Gilna, P.; Liao, S.  
Biochem. J. 257, 293-296, 1989  
A:Title: Identification of an androgen-repressed mRNA in rat ventral prostate as coding  
A:Reference number: S18491; MUID:89149740  
A:Accession: S18491  
A:Status: preliminary; nucleic acid sequence not shown; translation not shown  
A:Molecule type: mRNA  
A:Residues: 1-186, 'D', 188-447 <BET>  
A:Cross-references: EMBL:X13231; NID:g57240; PIDN:CAA31618.1; PID:g57241  
A:Note: The nucleotide sequence was submitted to the EMBL Data Library, October 1988  
R:Wong, P.; Pineault, J.; Lakins, J.; Taillefer, D.; Leger, J.; Wang, C.; Tenniswood, M.  
J. Biol. Chem. 268, 5021-5031, 1993  
A:Title: Genomic organization and expression of the rat TRPM-2 (clusterin) gene, a gene  
A:Reference number: A45415; MUID:93186813  
A:Accession: A45415  
A:Status: preliminary  
A:Molecule type: DNA; mRNA  
A:Residues: 1-447 <WON>  
A:Experimental source: blood  
A:Note: sequence inconsistent with the nucleotide translation  
A:Note: sequence extracted from NCBI backbone (NCBIN:126803, NCBI:P:126805)  
R:Cheng, C.Y.; Chen, C.L.C.; Feng, Z.M.; Marshall, A.; Bardin, C.W.  
Biochem. Biophys. Res. Commun. 155, 398-404, 1988  
A:Title: Rat clusterin isolated from primary Sertoli cell-enriched culture medium is sul  
A:Reference number: A90146; MUID:88326333  
A:Accession: B31575  
A:Molecule type: protein  
A:Residues: 22-51 <CHES>  
A:Note: amino end of the beta chain  
A:Accession: A31575  
A:Molecule type: protein  
A:Residues: 227-256 <CH2>  
A:Note: amino end of the alpha chain  
C:Superfamily: clusterin  
C:Keywords: apoptosis; disulfide bond; glycoprotein  
F:1-21/Domain: signal sequence #status predicted <SIG>  
F:22-226/Domain: clusterin beta chain #status experimental <BCH>  
F:227-447/Domain: clusterin alpha chain #status experimental <MAT>  
F:227-447/Domain: clusterin alpha chain #status experimental <ACH>  
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) (covalent) #status predicted

Query Match 14.4%; Score 380; DB 2; Length 447;  
Best Local Similarity 23.5%; Pred. No. 1.7e-16;  
Matches 113; Conservative 103; Mismatches 197; Indels 68; Gaps 15;

QY 34 LLVFIYCLLWLDKSHCAPTKWKTATSENLSKSFSEVEID-----ADEVKKALTGIKQ 87  
DB 4 LLLCVALL-----TWDMGVLGQEFSDNEIQLSTQGSRYRYEIONAVGVXKH 54  
QY 88 MKTMRKEKEHTNLMSTLKKCKREEKQKALKLNLEQVHELEERLCRESLADSWGECRS 147  
DB 55 IKTLIEKTNARKSLNLSLEAKKKEGALDDTDFDSEMLKAPFVCNETMMALWEECKP 114  
QY 148 CLENNCKRIYV-TCQPSWSSVKNKIERFFRKIYQVLPFHEDNEKOLPISEKLIKDAQL 206  
DB 115 CLKHTCKMFARVCRSGSLVGRQLEEFNLQSSPFYFWMNGDR-----IDSLLESDRQ 168  
QY 207 TQ-----MEDVFSQITVDVNSLFRNSNVFRMQOEEDQTFQSHFISDTDLTEPYFFP--- 259  
DB 169 SQVLDAWQDSFTRASGLIHFLFDQRF--FTHEPQDI-----HHF-----SPMGFPKRR 214  
QY 260 ---AFSKEPTKADLEQCWDIP--NFFOLFCNFSVIVESYSEITKMLKAIEDLPKQDKA 315

DB 215 PHFLYPKSRLVRSILMPLSHYGLPSLHNFQFFDMIHQAQQAMDVQLHSPALQFPDVFDFL 274  
QY 316 PDHGGSLISKMLPGQ--DRGLGCGELDNLSRCFKFHEKCKQKCOAHLSEDC-----PDVPALHT 370  
DB 275 KE-----GEDDPTVCKEIRHNSGCLKMKGCKEQEILSDVCSTNNPAQANLRQ 324  
QY 371 ELDEAIRLVNVSNOQYGOILQMTKRKHLLEDYALVKEKMRGQFGWVSELAN--QAPETELIF 428  
DB 325 ELNDSLAQVAERLTQOYNELLHSLQSKMLNTSSLLLEQLNDQFTWVSQLANLTQGDQDYLRV 384  
QY 429 NSIQVVPRIHEGNISKQDETMTDLSILPSSNFTLKIPLSESAESSNFIGVYVAKALQHF 488  
DB 385 STVTT-----HSSDEVPSRVTEVVVKLFSDSPITVWLPPEVSKDNPKFMDIVAELQLEY 440  
QY 489 K 489  
DB 441 R 441

RESULT 10  
148174  
sulfated glycoprotein 2 - golden hamster (fragment)  
C:Species: Mesocricetus auratus (golden hamster)  
C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 13-Aug-1999  
C:Accession: I48174  
R:Duguig, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989  
A:Title: Changes in brain gene expression shared by scrapie and Alzheimer disease.  
A:Reference number: I48174; MUID:89386721  
A:Accession: I48174  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-191 <RES>  
A:Cross-references: GB:M26640; NID:g191450; PIDN:AAA37102.1; PID:g191451  
C:Superfamily: clusterin

Query Match 6.4%; Score 168; DB 2; Length 191;  
Best Local Similarity 35.1%; Pred. No. 0.0012;  
Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps 1;

QY 330 DRGLGCGELDNLSRCFKFHEKCKQKCOAHLSEDC-----PDVPALHTDELDEATRLVNVSNQ 385  
DB 69 DRVCKEIRHNSGCLKMKGCKEQEILSDVCSTNNPAQAHRLQELNDSLQVAERLTOR 128  
QY 386 YCQILQMTKRKHLLEDYALVKEKMRGQFGWVSELAN 419  
DB 129 YNELLHSLQTKMLNTSSLLLEQLNEQFNWVSQLAN 162

RESULT 11  
TI4867  
Interaptin - slime mold (Dictyostelium discoideum)  
C:Species: Dictyostelium discoideum  
C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000  
C:Accession: TI4867  
R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.  
J. Cell Biol. 142, 735-750, 1998  
A:Title: Interaptin, an actin-binding protein of the alpha-actinin superfamily in Dic  
ts.  
A:Reference number: Z18248; MUID:98365468  
A:Accession: TI4867  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1738 <RIV>  
A:Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC34582.1  
C:Genetics:  
A:Gene: abpD  
A:Introns: 173/2; 1680/1

Query Match 6.2%; Score 164; DB 2; Length 1738;  
Best Local Similarity 21.1%; Pred. No. 0.029;

Qy	185	FHDNEKDLPISEKLIERDAQLQTOMEDVFSOLTVDVNS--LFNRSFNVFROMQQEFD---	239
Db	977	-----NQCKSLQETIVTTKDAAELDKLTKYISDYKTIEQMRLTNOKMNE-KSIQOGLSLE	1031
Qy	240	-----QTGFSHFTSDTLDP-	267
Db	1032	SLKRVKLRENTLSIDVSILKKQKEELSVLKGVOELTINNLEEKNYLNADVAKQLPKL	1091
Qy	268	KADLEQCWDIPNFOLFPCNFVSIVIYESVETTITMKLKAIEDLP-QKQAPDHGGLISKML	326
Db	1092	KKELESINDKQDLYQLQAATFKNELEAKVECLNNNTKTSKTULEKENEEKCQNLSASUKYI	1151
Qy	327	PGDGRGLCGELDQNL----SRCFKHEKCQKAHLSDCPDVPALHTDELDEARLVNVS	382
Db	1152	EIHENLLKVLDLENKKYYEGQLDDL-BGLKDVDTNFOELSKKHRTTFN	1204
Qy	383	NQ-----QYGQILOMQTRKHLEDLAYLEVKMARGFGWWSLANOAQPETEIFNISIQVPP	435
Db	1205	HESILLRQSASYKEKLSLASSENKOLSNKVSSITTKOVNELSPKASKVPEL----	1259
Qy	436	RRIHF-GNISQDETMTDLSILPSSNFTLIKIPLESAESSNFIFGVVAKALOHEK	489
Db	1260	LMHEYSQLGTFFDEKRRKALIASRNEEUR-SLKSELSKKRKLEVEYQKVLEEVR	1313
RESULT 13			
B70356			
Chromosome assembly protein homolog - Aquifex aeolicus			
C;Species: Aquifex aeolicus			
C;Date: 08-May-1998 #sequence_revision 08-May-1998 #text_change 02-Jun-2000			
C;Accession: B70356			
R;Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Gra			
V.			
Nature 392, 353-358, 1998			
A;Title: The complete genome of the hyperthermophilic bacterium Aquifex			
A;Reference number: A70300; MUID:98196666			
A;Accession: B70356			
A;Status: preliminary; nucleic acid sequence not shown; translation not sho			
A:Molecule type: DNA			
A;Residues: 1-1156 <AOF>			
A;Cross-references: GB:AE000699; NID:g2983238; PIDN:AAC06839.1; PID:g298323			
A;Experimental source: strain VF5			
C;Genetics:			
A;Gene: xcpC			
C;Superfamily: chromosome segregation protein SMC1			
Query Match 5.7%; Score 150; DB 2; Length 1156;			
Best Local Similarity 20.3%; Pred.No. 0.14;			
Matches 90; Conservative 88; Mismatches 162; Indels 104; Gaps			
Qy	63	LKPSEVGIED----ADEEVKALTGIQIMIMP-----RKEKETNLMSTLKK	108
Db	163	IIEISIGIGEYERKKEALELAVELKIREDLIIEISNQLKRLKEEKLEKFELQR	222
Qy	109	CREEKBQALKILLEVBHEEREIRCRE--SLADSWGECRCSELNNCMRYITTCOPSWSS	166
Db	223	IKRE-TRAKITLKEKELLKERERIINLSURESLEDITFIQBNEXEL-----	271
Qy	167	VKNIERFFFRKYQFLPFPHEDNEK---DLPTISEKILI-EKDAQLTQMEDVFSQLTVDVNS	222
Db	272	--NERERLLKEVNEMIPPFKEVGYTAELIENAERSIKEKERELNESNRVKNLEELINN	329
Qy	223	LFNRSFNVRQ---MQEQFDQTFOSHIFSDTDITEPY-FFPAFSKPEPMTKADLECWDIP	278
Db	330	LLSDPENLEREVGTLQLELEK-----LKEEYKSLKEVEREKLRELEEE----	373
Qy	279	NFFOLF CNF-VSVSYIESVSETITMKLKAIEDLPQDKAPDHGGILLSKMLPGODRGLCGEL	337
Db	374	--ERLKITTFDEVKKLEBEKEKLTTEKUNSL-NKEQELETORANLNKKI-----BR	421
Qy	338	DONLSRCFCFKPEKCKQCAHLSDCPDVPALHTDELDEARLVNVSNOQYGOILOMQTRKHL	397

Db 422 KEDINKLISREKIKKEKEIKRLKAIAKKKEEELRNLTQELNIYEKRLSEVRKKL 481  
QY 398 E-----DTAYLVKMRGQFGWVSELAN-QAPE-----423  
Db 482 EYLVKEKGATEREVRFSVDYDFDKIKGVYGSVELIRVKNPEHITAIEVAGGRLKFI 541  
QY 424 ----TEIIFNSIQVVPRIHEGNIS 443  
Db 542 VVEDEEVAKEICQLAKRMNLGRFS 565

## RESULT 14

S70633  
serine/threonine-specific protein kinase (EC 2.7.1.1); Rho-associated - bovine  
N:Alternate names: Rho-associated protein kinase  
C:Species: Bos primigenius taurus (cattle)  
C:Date: 14-Feb-1997 #sequence\_revision 13-Mar-1997 #text\_change 24-Sep-1999  
C:Accession: S70633; S77694  
R:Matsumi, T.; Anano, M.; Yamamoto, T.; Chihara, K.; Nakafuku, M.; Ito, M.; Nakano, T.; O  
EMBO J. 15, 2208-2216, 1996  
A:Title: Rho-associated kinase, a novel serine/threonine kinase, as a putative target fo  
A:Reference number: S70633; PMID:96208507  
A:Accession: S70633  
A:Status: nucleic acid sequence not shown  
A:Molecule type: mRNA  
A:Residues: 1-1388 <MAT1>  
A:Cross-references: EMBL:U36909; NID:g1326077; PIDN:AA048567.1; PTD:g1326078  
A:Accession: S77694  
A:Molecule type: protein  
A:Residues: 1-1070 <MAT2>  
C:Superfamily: unassigned Ser/Thr or Tyr-specific protein kinases; protein kinase C zinc  
C:Keywords: ATP; phosphotransferase; serine/threonine-specific protein kinase  
F:90-354/Domain: protein kinase homology <KIN>  
F:98-106/Region: protein kinase ATP-binding motif  
F:1261-1315/Domain: protein kinase C zinc-binding repeat homology <KZN>

Query Match 5.7%; Score 149; DB 2; Length 1388;  
Best Local Similarity 16.7%; Pred. No. 0.2;  
Matches 95; Conservative 96; Mismatches 166; Indels 212; Gaps 20;  
QY 54 KDKTAISENLKSFSEVGEIDAEYKALTGIKOMKI-----MMERKEKEHTNLM 104  
Db 780 KQKQVLNEDVRNLT-----LKIEQETQKRLTQNDLKMOTQVNTLKMSEKQLKQENHLL 835  
QY 105 TLK-----KCREKQKALLLNEVQHLDEE-----RLCR 135  
Db 836 EMKMSLEKQNAELRKERQADQGMKELQDLAEQYFTLYKTQVRELKECECEKTKLCK 895  
QY 136 E-----SLADSWGCRSCLNENCRIVTTCOPSSVKNKIERFRKIYQFLPFPHEDN 189  
Db 896 ELQKKQKQELQDE-----RSLAAQLEITLTKADSEQARSIAEEQYSDLEKE 942  
QY 190 E--KDLPISE-----KLIKEDAQLTQMEDVFSOLTVDVNSLNFNSFVRMQOEFDDQ 240  
Db 943 KIMKELEIKEMMARHQELTEKDATIASLEETNRLTSDVANLANEKEELNNKLEAQEQ 1002  
QY 241 TQSHFISDTDLTEYFFPAPSKPEMTKADLEQCWDIPNFQFCNFSVYESVSETIT 300  
Db 1003 LSR-----LKDEEISAAAIKAQFEKQLLTERTLK-----TQAVN 1036  
QY 301 KMLKAIE-----DLPKQKAPDHGGLISKMLPGQDRGLCGEL-----DQNLRCF 345  
Db 1037 KLAETMNRKEPVKRGNDTIVRKEK-----ENKRLHMLKLSREKLTQOMI 1082  
QY 346 KPFHEKQCKQAHLSBDCPDVPALHTDELAIPLVNSVNSQYQGLQMTKRKHLE-----398  
Db 1083 RYQKELNEMOQIAIE-----SQIRIELQMTLDSKDSIEQLRSOLQALHIGLSSSIGG 1138  
QY 399 -----DTAYLVKMRG-----QFGVSEL-----417  
Db 1139 PGDTAADGFPESRLEGLWLSLVRNNTKKFGWKVYIVSSKKILFLFYDSEQDKQSNPYM 1198

QY 418 -----ANOAPETEIIFNISIQVVPRI-----HEGNISKODETMTDLSILPSSNFT 462  
Db 1199 VLDIDKLFHVRVPTQDYYRADAKEIPRFQILYANEGESKKEQE-----1243  
QY 463 LKIPLEESAESSNFI---GYVVAKALQHF 488  
Db 1244 --FPVEPVGEKSNYICKHGHEFIPTLYHF 1270  
RESULT 15  
T34107  
hypothetical protein C18C4.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34107  
R:Gattung, S.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid C18C4.  
A:Reference number: Z21478  
A:Accession: T34107  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-1091 <GAT>  
A:Cross-references: EMBL:U55369; PIDN:AAC25825.1; GSPDB:GN00023; CESP:C18C4.5  
A:Experimental source: strain Bristol N2; clone C18C4  
C:Genetics:  
A:Gene: CESP:C18C4.5  
A:Map position: 5  
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2;  
Query Match 5.6%; Score 148.5; DB 2; Length 1091;  
Best Local Similarity 20.3%; Pred. No. 0.16;  
Matches 104; Conservative 100; Mismatches 184; Indels 125; Gaps 25;  
QY 54 KDKTAISENLKSFSEVGEID-----ADEVKKALTGIKOMK--IMMERKEKEHTNLM 103  
Db 283 KDNNAIL-NVOLREKDGKIDRIQVDLLAAESRAQAQAEEDVRDMKERIITSKDDSNLL 341  
QY 104 -STLKKCKREEKQKALKLNEVQHLDEEERLCRESLADSWGCRSCLNENCMR-----IY 157  
Db 342 QDELRLRTTEKQQQAQKKIENDETIKQETQIRD-LGRSLDEAKKQLQKMQSEQRNEEVA 400  
QY 158 TTCQPSWSVSNKIERFFRKIYQFLPFPHEDNEKDLPISEKLIKEDAQLTQ-----208  
Db 401 RQGEDSARSMBEEKATKEEIKKLKSOVLQQLQEDLEQLKKRVQ---ELTEQKRVLESKA 457  
QY 209 -MEDVFSOLTVDVNSLNFNSFVRMQOEF-----DQTFQ-----243  
Db 458 SVADEFGTLMSSLSL--REEN--RQYEEETRSLQTNIRTLQDEVYQHQAITEMKNRAE 513  
QY 244 -----SHFIS-----DTDLTEPYFFPAPSKPEMTKADLEQCWDIPNFQFCNFS 288  
Db 514 KAEVIEKENHRVQNASSSHDADITRLENEXTQMBEALAKADQEKQAIRE-----ASES 568  
QY 289 VSIYE-----SVSETITKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLCGELQNLNRC 344  
Db 569 VRVKRMTEASITSDRIOQLSKE--KVDS-----LTRELESSRRRM-EQLQEDQTKF 618  
QY 345 KPFHEKQCKQAHLSBDCPDVPALHTDELDEAIRLVNSVNSQYQGLQMTKRKHLEDTAYLV 404  
Db 619 LGSHD---ETKAEWMKD-----LHEAQDETEKLTNQAOLKSKNETLTTE-LEDSONLC 668  
QY 405 EKMKGQFGWV----SELANQAPETEIIFNISIQVVPRIHEGNISKODETMTDLSILPSSN 460  
Db 669 ERLKAQEKADKDYEEKTVQLREAEDLADRLQAA-QILSGNVESKFSQMKQE-----719  
QY 461 FTLKIPLEESAESSNFIQYVVAKALQHFKEHFK 493  
Db 720 --SKIEMERILDNHN-----KELEKIRELK 743

Search completed: July 2, 2002, 11:52:00  
Job time: 223 sec

---

---

**THIS PAGE BLANK (USPTO)**



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:43:54 ; Search time 24 Seconds  
(without alignments)  
485.459 Million cell updates/sec

Title: US-09-722-544A-4  
Perfect score: 2529  
Sequence: 1 MRTWDSYSGNMKPLLVF.....FIGYVAKALQHFKEHFTW 477

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep.\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep.\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep.\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep.\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep.\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2487	98.3	495	US-09-268-992-2	Sequence 2, Appli
2	2464	97.4	466	US-09-268-992-64	Sequence 64, Appli
3	2377	94.0	477	US-09-268-992-4	Sequence 4, Appli
4	2350	92.9	446	US-09-268-992-51	Sequence 51, Appli
5	1755.5	69.4	465	US-09-268-992-49	Sequence 49, Appli
6	1639	64.8	466	US-09-268-992-39	Sequence 39, Appli
7	1558	61.6	450	US-09-268-992-41	Sequence 41, Appli
8	1277	50.5	374	US-09-268-992-43	Sequence 43, Appli
9	1272.5	50.3	373	US-09-268-992-45	Sequence 45, Appli
10	776	30.7	208	US-09-268-992-67	Sequence 67, Appli
11	145	5.7	1388	US-08-685-576-1	Sequence 1, Appli
12	141	5.6	1388	US-08-685-576-4	Sequence 4, Appli
13	130.5	5.2	2482	US-08-328-254-6	Sequence 6, Appli
14	128.5	5.1	3248	US-08-353-700-1	Sequence 1, Appli
15	128.5	5.1	3248	PCT-US95-16216-1	Sequence 1, Appli
16	125.5	5.0	1354	US-08-685-871-2	Sequence 2, Appli
17	121	4.8	828	US-08-993-228-21	Sequence 21, Appli
18	119.5	4.7	3111	US-08-460-309-4	Sequence 4, Appli
19	119.5	4.7	3111	US-08-125-077-4	Sequence 4, Appli
20	118.5	4.7	1618	US-07-853-913-4	Sequence 4, Appli
21	118	4.7	1886	US-08-938-105-3	Sequence 3, Appli
22	118	4.7	2154	US-08-841-349-4	Sequence 4, Appli
23	117	4.6	1098	US-08-923-992A-8	Sequence 8, Appli
24	117	4.6	1104	US-08-923-992A-4	Sequence 4, Appli
25	117	4.6	1939	US-09-310-187A-1	Sequence 1, Appli
26	117	4.6	2052	US-09-045-201A-2	Sequence 2, Appli
27	116.5	4.6	816	US-08-533-306A-6	Sequence 6, Appli

28	116.5	4.6	816	2	US-08-742-923A-6	Sequence 6, Appli
29	116.5	4.6	1057	4	US-09-541-782-10	Sequence 10, Appli
30	116	4.6	1164	4	US-08-923-992A-2	Sequence 2, Appli
31	115.5	4.6	885	2	US-08-533-306A-4	Sequence 4, Appli
32	115.5	4.6	885	2	US-08-742-923A-4	Sequence 4, Appli
33	115.5	4.6	1786	4	US-08-973-462-8	Sequence 8, Appli
34	114.5	4.5	452	2	US-08-686-599A-18	Sequence 18, Appli
35	114.5	4.5	493	2	US-08-686-599A-5	Sequence 5, Appli
36	114.5	4.5	493	2	US-08-686-599A-16	Sequence 16, Appli
37	114	4.5	435	2	US-08-531-439B-4	Sequence 4, Appli
38	114	4.5	1128	4	US-08-923-992A-6	Sequence 6, Appli
39	112.5	4.4	467	2	US-08-686-599A-17	Sequence 17, Appli
40	110.5	4.4	1066	4	US-09-541-782-8	Sequence 8, Appli
41	110.5	4.4	1184	4	US-09-541-782-2	Sequence 2, Appli
42	110	4.3	976	4	US-09-104-324B-4	Sequence 4, Appli
43	108	4.3	2285	4	US-09-308-375-2	Sequence 2, Appli
44	107	4.2	1164	4	US-08-923-992A-10	Sequence 10, Appli
45	107	4.2	1497	1	US-08-623-679-7	Sequence 7, Appli

ALIGNMENTS

RESULT 1  
US-09-268-992-2  
; Sequence 2, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; EARLIER FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 2  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-2

Query Match	98.3%	Score	2487	DB	4	Length	495
Best Local Similarity	99.8%	Pred. No.	2.6e-229				
Matches	470	Conservative	1	Mismatches	0	Indels	0
Qy	7	SNSGNMKPPLLVFVLCVLLKDSHCAPTWKDKTAISENLKSFSEVGETDADEEVKALTG	66				
Db	25	NSGNMKPPLLVFVLCVLLKDSHCAPTWKDKTAISENLKSFSEVGETDADEEVKALTG	84				
Qy	67	IKQKIMMERKEKHTNLMSTLKKREEKQKALKEVQHLKEEERLCRESLADSWGE	126				
Db	85	IKQKIMMERKEKHTNLMSTLKKREEKQKALKEVQHLKEEERLCRESLADSWGE	144				
Qy	127	CRSCLNNCRITYTCOPSSVSKNKIERFRKTYQFLFPFHEDNEKDLPISEKLIBEDA	186				
Db	145	CRSCLNNCRITYTCOPSSVSKNKIERFRKTYQFLFPFHEDNEKDLPISEKLIBEDA	204				
Qy	187	OLTQMEDVFQSLTVDVNSLNRSENFVROMQOEFQDFQSHFISDTLTTEYFFPAFSKE	246				
Db	205	OLTQMEDVFQSLTVDVNSLNRSENFVROMQOEFQDFQSHFISDTLTTEYFFPAFSKE	264				
Qy	247	PMTKADLEQCWDIPNFFQLFCNFSVSVSVSETITKMLKAIEDLPKQDKAPDHGLISK	306				

Db 265 PMTKADLEQCDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISK 324  
Qy 307 MLPQDGRGLCGELDQNLSCRFKHEKQKQAHLSDCDPDPALHTLDEAIRLVNVSQ 366  
Db 325 MLPQDGRGLCGELDQNLSCRFKHEKQKQAHLSDCDPDPALHTLDEAIRLVNVSQ 384  
Qy 367 QYQIQLQMTKRKLEDATYLVKMRGQFGWSELANOAPETEIFNSIQVVPRIHEGNTSK 426  
Db 385 QYQIQLQMTKRKLEDATYLVKMRGQFGWSELANOAPETEIFNSIQVVPRIHEGNTSK 444  
Qy 427 QDETMTDLSILPSSNFTLKIPLEESAESSNFIGYVAKALOHEKHEFKTW 477  
Db 445 QDETMTDLSILPSSNFTLKIPLEESAESSNFIGYVAKALOHEKHEFKTW 495

RESULT 2  
US-09-268-992-64  
; Sequence 64, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 64  
; LENGTH: 466  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-64

Query Match 97.4%; Score 2464; DB 4; Length 466;  
Best Local Similarity 100.0%; Pred. No. 3.7e-227;  
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 MKPPLLVFIVCLLWKDSCAPTWKDKTAISENLKSFSEVGEDADEEVKKALTGIKQMK 71  
Db 1 MKPPLLVFIVCLLWKDSCAPTWKDKTAISENLKSFSEVGEDADEEVKKALTGIKQMK 60  
Qy 72 IMMERKEKHTNLMSTLKKCKREEKQKALKLLNEVOEHLEEEERLCRESLADSWGCRSL 131  
Db 61 IMMERKEKHTNLMSTLKKCKREEKQKALKLLNEVOEHLEEEERLCRESLADSWGCRSL 120  
Qy 132 ENNCRIYTTQPSWSSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLIBEDAQLTOM 191  
Db 121 ENNCRIYTTQPSWSSVKNKIERFRKIYQFLFFPHEDNEKDLPISEKLIBEDAQLTOM 180  
Qy 192 EDVFSQLTVDVNSLNRSNFVRMQOQEDDTQFQSHFISDITLTPYFPFPAFSKEPMTKA 251  
Db 181 EDVFSQLTVDVNSLNRSNFVRMQOQEDDTQFQSHFISDITLTPYFPFPAFSKEPMTKA 240  
Qy 252 DLEQCDWDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPCQ 311  
Db 241 DLEQCDWDIPNFQFCNFSVSIYESVSETITKMLKAIEDLPKQDKAPDHGGLISKMLPCQ 300  
Qy 312 DRGLCGELDQNLSCRFKHEKQKQAHLSDCDPDPALHTLDEAIRLVNVSNOOYGOI 371  
Db 301 DRGLCGELDQNLSCRFKHEKQKQAHLSDCDPDPALHTLDEAIRLVNVSNOOYGOI 360

Qy 372 LQMTKRKHELDATYLVKMRGQFGWSELANOAPETEIFNSIQVVPRIHEGNTSKODETM 431  
Db 361 LQMTKRKHELDATYLVKMRGQFGWSELANOAPETEIFNSIQVVPRIHEGNTSKODETM 420  
Qy 432 MTDLSTLSPSSNFTLKIPLEESAESSNFIGYVAKALOHEKHEFKTW 477  
Db 421 MTDLSTLSPSSNFTLKIPLEESAESSNFIGYVAKALOHEKHEFKTW 466

RESULT 3  
US-09-268-992-4  
; Sequence 4, Application US/09268992  
; Patent No. 6342351  
; GENERAL INFORMATION:  
; APPLICANT: Chen, H.  
; APPLICANT: Freimer, N.  
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING  
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS  
; FILE REFERENCE: 7853-138  
; CURRENT APPLICATION NUMBER: US/09/268,992  
; CURRENT FILING DATE: 1999-03-16  
; EARLIER APPLICATION NUMBER: 09/236,134  
; EARLIER FILING DATE: 1999-01-22  
; EARLIER APPLICATION NUMBER: 60/106,056  
; EARLIER FILING DATE: 1998-10-28  
; EARLIER APPLICATION NUMBER: 60/088,312  
; EARLIER FILING DATE: 1998-06-05  
; EARLIER APPLICATION NUMBER: 60/078,044  
; EARLIER FILING DATE: 1998-03-16  
; NUMBER OF SEQ ID NOS: 84  
; SOFTWARE: FastSeq for Windows Version 3.0  
; SEQ ID NO 4  
; LENGTH: 477  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-268-992-4

Query Match 94.0%; Score 2377; DB 4; Length 477;  
Best Local Similarity 91.5%; Pred. No. 8e-219;  
Matches 455; Conservative 1; Mismatches 1; Indels 40; Gaps 2;

Qy 1 MRETWYSNGNKKPPLLVFIVCLLWKDSCAPTWKDKTAISENLKSFSEVGEDADEEV 60  
Db 1 MRETWYSNGNKKPPLLVFIVCLLWKDSCAPTWKDKTAISENLKSFSEVGEDADEEV 60  
Qy 61 KKALTGIKQKIMMERKER-----EHTNLMSTLKKCKREEKQKALK 100  
Db 61 KKALTGIKQKIMMERKERKANQAPETEIFNSIQVVPRIEHTNLMSTLKKCKREEKQKALK 120  
Qy 101 LLNEVOEHLEEEERLCRESLADSWGCRSCLNNCMRIYTTQPSWSSVKNKIERFRKI 160  
Db 121 LLNEVOEHLEEEERLCRESLADSWGCRSCLNNCMRIYTTQPSWSSVKNKIERFRKI 180  
Qy 161 YOFLPFPHEDNEKDLPISEKLIBEDAQLTOMEDVFSQLTVDVNSLNRSNFVRMQOQEF 220  
Db 181 YOFLPFPHEDNEKDLPISEKLIBEDAQLTOMEDVFSQLTVDVNSLNRSNFVRMQOQEF 240  
Qy 221 DQTFQSHFISDITLTPYFPFPAFSKEPMTKADLEQCDWDIPNFQFCNFSVSIYESVSET 280  
Db 241 DQTFQSHFISDITLTPYFPFPAFSKEPMTKADLEQCDWDIPNFQFCNFSVSIYESVSET 300  
Qy 281 ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDGRGLCGELDQNLSCRFKHEKQKQAHL 340  
Db 301 ITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDGRGLCGELDQNLSCRFKHEKQKQAHL 360  
Qy 341 SEDCPDVPALHTLDEAIRLVNVSNOOYGOIILQMTKRKHELDATYLVKMRGQFGWSEL 400  
Db 361 SEDCPDVPALHTLDEAIRLVNVSNOOYGOIILQMTKRKHELDATYLVKMRGQFGWSEL 419  
Qy 401 NOAPETEIFNSIQVVPRIHEGNTSKODETMTDLSILPSSNFTLKIPLEESAESSNFITG 460  
Db 420 -----HEGNTSKODETMTDLSILPSSNFTLKIPLEESAESSNFITG 460





```

; SEQ ID NO 43
; LENGTH: 374
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-43

```

Query Match 50.5%; Score 1277; DB 4; Length 374;  
Best Local Similarity 54.5%; Pred. No. 7.5e-114;  
Matches 255; Conservative 44; Mismatches 73; Indels 9

Qy	12	MKPPLLVFI	VCLLWLK	DSCAPTWK	OKTALISE	NKLSFSE	VEGEIDA	DEEVKKALT	GIGKQM	71
Db	1	1	1	1	1	1	1	1	1	1
Qy	72	IMMERKEHT	NLMSTL	LKKCKREK	QALKLL	NVQVSH	LEEREL	CRESLAD	SGCRSL	131
Db	61	1	1	1	1	1	1	1	1	1
Qy	132	ENCMRIY	TTCQPS	WSWSN	KNIERR	FRKIYQ	FLFP	PHEDNE	KDLP	191
Db	121	ESNCRM	FDTT	TCQPAW	SSVK-					140
Qy	192	EDVFSQ	LTVDN	SLFN	RSENV	FRQM	QOE	FDOT	FOQSH	251
Db	141	1	1	1	1	1	1	1	1	1
Qy	252	DLEQCWD	IPNF	FOLF	NFSYI	ESYSE	ITIK	MLKA	IEDLP	311
Db	148	DAEPSW	AI	PVNF	L	CLN	LSF	VSQV	SEKLT	207
Qy	312	DRGLCEL	DONLS	RCP	FEK	CKO	CAH	LSDEC	DPV	371
Db	208	DRGSDG	LKGN	LSDC	VNFR	KRCQ	KODY	LSDDC	PVNP	267
Qy	372	LQMTKR	HLED	TAYL	LV	EKM	RQFC	WGS	ELAN	429
Db	268	QVMTQY	HLED	TTL	LM	EKM	RQFC	WGS	ELAN	327
Qy	430	TMTD	SILP	SSNFT	L	KIP	LES	AE	SSNFI	477
Db	328	TVVPS-	SLLP	SSNFT	LS	SPLEK	SAG	NANF	IDHV	374

```

RESULT          9
US-09-268-992-45
; Sequence 45, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freilmer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 45
; LENGTH: 373
; TYPE: PRT
; ORGANISM: Cavia sp.
US-09-268-992-45

```

Query Match 50.3%; Score 1272.5; DB 4; Length 373;  
Best Local Similarity 54.5%; Pred. No. 2e-113;  
Matches 255; Conservative 45; Mismatches 71; Indels 97;

Qy	12	MKPPLLVFIVCLLWKDSCAPTWKOKTATISENLKSFSEYGEIDADEEVKKALTGIKOMK	71
Db	1	1 KMLPLFPVCLLWLKDCHCAPTWKOKTATISENANSFSEAGEIDVGEVKIALIGIKOMK	60
Qy	72	IMMERKEHTNLWSTLTKCKEEOKALKLLNEVOEHLEEEERLCRESLADSHGECRSL	131
Db	61	IMMERREEHSHKMLTKCKEEOKALKUMNEVHEHLEEEELCSQVSLADSWDECRAL	120
Qy	132	ENCMCRITVQPCSWSSWNKKIERFRKIKYQFLPPFHEONEKDLPISEKLIBEDAUTQM	191
Db	121	ESNCMRDFTTCQPAWSSVK-	140
Qy	192	EDVFSQLTVDVNSLFRNSFNVMQOEFTQTSHFISDTEPYFPFAPSKPEMTKA	251
Db	141	-----RA 146	
Qy	252	DLEQCWDIPNFFQLCNFSVSIYESVSETITKMLKAIEDLPKODKADPHGGGLISKMLPGQ	311
Db	147	DAEPSWAIPNVFOLLNLSFVSQVSEKLIITLTRATEDPPKODKSNQGGPISKILPEQ	206
Qy	312	DRGLCGELDONLSRCPKFHEKCKOKQAHLSDECDPVPALHTELDEAIRLVNYSNOQYGOI	371
Db	207	DRGSDGLGONLSDCVNFRKCRKOCODYLSDDCPNPPELYREILNEAIRLVNYSNOQYDQV	266
Qy	372	LQMKRHELDATYLVKMRQFGWSELANOAPETIEFINSIQVPR1--HEGNISKQDE	429
Db	267	VQMTQYHLEDTLLMEKMRQFGWSELAYQSGAEDIFNPVKWVALSAHEGNSDQDD	326
Qy	430	TMWTDLSILPSSNFTKIPILBESAESSNFTIGYVVAKALQHFKEHFKTW	477
Db	327	TVVPS--SLPSSNFTSSLPLEKSAGANAFTHDVVEKVLQHFKEHFKHTW	373

```

RESULT 10
US-09-268-992-67
; Sequence 67, Application US/09268992
; Patent No. 6342351
; GENERAL INFORMATION:
; APPLICANT: Chen, H.
; APPLICANT: Freimer, N.
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR DIAGNOSING
; TITLE OF INVENTION: AND TREATING CHROMOSOME-18p RELATED DISORDERS
; FILE REFERENCE: 7853-138
; CURRENT APPLICATION NUMBER: US/09/268,992
; CURRENT FILING DATE: 1999-03-16
; EARLIER APPLICATION NUMBER: 09/236,134
; EARLIER FILING DATE: 1999-01-22
; EARLIER APPLICATION NUMBER: 60/106,056
; EARLIER FILING DATE: 1998-10-28
; EARLIER APPLICATION NUMBER: 60/088,312
; EARLIER FILING DATE: 1998-06-05
; EARLIER APPLICATION NUMBER: 60/078,044
; EARLIER FILING DATE: 1998-03-16
; NUMBER OF SEQ ID NOS: 84
; SOFTWARE: FastSeq for Windows version 3.0
; SEQ ID NO 67
; LENGTH: 208
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-268-992-67

```

Query Match 30.7% Score 776; DB 4; Length 208;  
Best Local Similarity 98.6%; Pred. No. 2.4e-66;  
Matches 145; Conservative 2; Mismatches 0; Indels 0; Gaps 0;



ATTORNEY/AGENT INFORMATION:  
 NAME: Bent, Stephen A.  
 REGISTRATION NUMBER: 29,768  
 REFERENCE/DOCKET NUMBER: 16887/843  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (202)672-5300  
 TELEFAX: (202)672-5399  
 TELEX: 904136  
 INFORMATION FOR SEQ ID NO: 4:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 1388 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-685-576-4

Query Match 5.68; Score 141; DB 2; Length 1388;  
 Best Local Similarity 16.88; Pred. No. 0.00021;  
 Matches 94; Conservative 97; Mismatches 176; Indels 192; Gaps 20;

36 KDKTAISENLKSFSEVGEIDAEEVKKALTGKIKOMKI-----MMERKEKEHTNLS 86  
 780 KOKVLDNEVRNL-----LKIEQETQRCLTQNDLKWTQOVNTLKMSEKQLKQENHLM 835  
 87 TLK-----KREKEQKALKLLNEVOHLEEE-----RLCRESLADSWGEC-----127  
 836 EMKNLEKQNAELRKERQDADGOMKELQDLQAEQYFSTLYKTQVRELKECEKTKLKG 895  
 128 ---RSCLENNCMRIYTCQPSWSSVKKIERFFRKIYQFLPPFHEDNE--KDLPISE---179  
 896 ELQOKKQELQDERSLAQAEITITKADSEQLARSIAEQYSLEKEKIMKELEIKEMMA 955  
 180 ----KLIEEDAQLTQMEDVFSQLTVDVNSLFRNS---FNVRMOQOEFDOTFQSHETSDT 232  
 956 RHKQELTEKDATIASLEETNRTLSDVANLANEKEELNKKLDVQEQLSR-----LKDE 1009  
 233 DLTPYFPFAPSKPMYKADLEOCWDIPNFFQLFCNFSVSIYESVETITKMLKAIE---289  
 1010 EISAAAIKAQFQKLLERTLK-----TQAVNKLAEIMNRKE 1046  
 290 -----DLPKQKAPDHGGLSKMLPGQDRGLCGEL-----DONLSRCFKFHEKCKQCK 337  
 1047 PVKRGNTDVRKEK-----ENRKLHMLKEREKLTQOMIKYOKELNEMQ 1092  
 338 AHLSEDCDVPALHTELDEAIRLVNSNQYGOILQMTKRHLE-----DTAY 384  
 1093 AQIAEE---SQRIELQMTLDSKSDIEQLRSQALHGLDSSSTSGSGPDAEADGGF 1148  
 385 LVEKMRG-----QFGWSEL-----A 400  
 1149 PESLEGWLSLVRNNTKFGWKYVIVSVSKILFYDSEQDKREQSNPYMVLIDIKLFHV 1208  
 401 NQAPETETIIFNSIQVPR-----HKGNTSKODETMTDLSILPSSNFTKIPLESAAE 454  
 1209 RPVTQTDVYRADAKEIPRIQILYANEGESKKEOE-----FPVEPVGE 1251  
 455 SSNFI---GYVAKALQHF 470  
 1252 KSNYICHKGHEFTLYHF 1270

RESULT 13  
 US-08-328-254-6  
 Sequence 6, Application US/08328254  
 Patent No. 5710022  
 GENERAL INFORMATION:  
 APPLICANT: Zhu, Xuelliang  
 TITLE OF INVENTION: A No. 5710022el Nuclear Mitotic Phosphoprotein  
 NUMBER OF SEQUENCES: 8  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Campbell and Flores

STREET: 4370 La Jolla Village Drive, Suite 700  
 CITY: San Diego  
 STATE: California  
 COUNTRY: USA  
 ZIP: 92122  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/328,254  
 FILING DATE: 24-OCT-1994  
 CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/141,239  
 FILING DATE: 22-OCT-1993  
 ATTORNEY/AGENT INFORMATION:  
 NAME: Campbell, Cathryn A.  
 REGISTRATION NUMBER: 31,815  
 REFERENCE/DOCKET NUMBER: P-CJ 1191  
 TELECOMMUNICATION INFORMATION:  
 TELEPHONE: (619) 535-9001  
 TELEFAX: (619) 535-8949  
 INFORMATION FOR SEQ ID NO: 6:  
 SEQUENCE CHARACTERISTICS:  
 LENGTH: 2482 amino acids  
 TYPE: amino acid  
 TOPOLOGY: linear  
 MOLECULE TYPE: protein  
 US-08-328-254-6

Query Match 5.28; Score 130.5; DB 1; Length 2482;  
 Best Local Similarity 18.88; Pred. No. 0.0051;  
 Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

27 KDSHCAPTMDKDTAISNLK-----SFSEVGE-----IDADE---EVKRALTGIKQ 69  
 1502 KQLHIAEKLERENDSLADKVENLERELQMSSENOELVILDAENSKAEVETIKTQIEE 1561  
 70 M-----KIMMERKEKEH-----TNLMSTLKKCREKQKALKLNEV 105  
 1562 MARSKLVFELDLVTLRSEKENLTKIQKQGLSELKLLSSFKSLLEKEQA-----EI 1616  
 106 QEHLSEERLCRESLADSWGECRSCLENNC--MRIYTCQPSWS-----SVKNKIERF 156  
 1617 Q--IKESKTAVEMQLQKELNEAVALCGDQEIIMKATEQSLDPPIEEHQLRNSIEKL 1674  
 157 FRKIYQFLFPFHEDNEKDLPISSEKLIBED-----185  
 1675 RARL-----EADKKQLCVLQQLKESEHHDLLKGRVENLERELIARTNOEHALEA 1727  
 186 -----AQLTQMEDVFSQLTVDVNSLFRNSFNVRMOQOEFDOTFQSHFISTDL 234  
 1728 ENSKGEVETLKAKIEGTQSLRGLDVLVIRSEKENTLNELOKEQER-----ISELEI 1781  
 235 TEPYFFAPSKPEMTKADLEOCWDIPNFFQLFCNFSVSIYESVETITKMLKA--IEDLPK 293  
 1782 INSSFENILOEKEQEKVQMAE-----KSSNAMEMLOTQLKELNE 1820  
 294 QDKAPDHGGLSKMLPGQDRGLCGELDQNLIS---RCFKFHEKCKQKQAHLSDCPDVPAL 350  
 1821 RVAALHN-----DQEAACKAKEQNLSSQVECELEL-EKAQLLOG-----1856  
 351 HTLDELDAIRLVNSNQYGOILQMTKRHLEDYALVEKMRGQFGWGSSELANQAPETETIIF 410  
 1857 ---LDEAKNNYIVLQSSVNGLIQ-----EVEDGKQKLEKDEE---ISRLKNQIQDQEQLV 1906  
 411 NSTQVVPRIH-----EGNISKQDETMTD--LSILPSSNFTKIPLE 450  
 1907 SKLSQVEGEHQLWKQENLRLNLTVELEQKIQVLSKNASLODTLE 1952

RESULT 14  
US-08-353-700-1  
; Sequence 1, Application US/08353700  
; Patent No. 5599919  
; GENERAL INFORMATION:  
; APPLICANT: YEN, TIMOTHY J.  
; TITLE OF INVENTION: NUCLEIC ACID ENCODING A  
; TITLE OF INVENTION: TRANSIENTLY-EXPRESSED KINETOCHORE PROTEIN,  
; TITLE OF INVENTION: AND METHODS OF USE  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: DANN, DORFMAN, HERRELL AND SKILLMAN  
; STREET: 1601 MARKET STREET, SUITE 720  
; CITY: PHILADELPHIA  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/353,700  
; FILING DATE: 09-DEC-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: REED, JANET E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; ORIGINAL SOURCE:  
; ORGANISM: HUMAN  
US-08-353-700-1

Query Match 5.1%; Score 128.5; DB 1; Length 3248;  
Best Local Similarity 18.8%; Pred. No. 0.012;  
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

QY 27 KUSHCAPTWKDKTAISENLK-----SFSEVGE-----IDADE---EVKKALGTGIQ 69  
DB 2230 KQLHIAEKLERERENDSKDKVENLERELQSEENQELVILDAENSKAEVETLKTQIEE 2289  
QY 70 M-----KIMMERKEKEH-----TNLMSTLKKCKREKQKALKLLNEV 105  
DB 2290 MARSLKIFDLVTLRSEKENLTKQIEKQGLSELDKLLSSPKSLLEEKEQA-----EI 2344  
QY 106 QEHLBEERLCRESLADSGECRSCLENNC--WRITTCQPSWS-----SVKNKIERF 156  
DB 2345 Q---IKESRTAVEMLQNLQKELNEAVALCGOQEIIMKATEQSLDPPIEBEHQLRNSIEKL 2402  
QY 157 FRKIYQFLPFFHDNKKLPISKELTIED-----SVKNKIERF 185  
DB 2403 RARL-----EADKQKLCVLOQLKESEHHDLLKGRVENLERELEIARTNOEHALEA 2455  
QY 186 -----AQLTQMEDVFSOLTVDVNSLFRNSFNVRQMQQEPDQTFQSHFISDTDL 234  
DB 2456 ENSKGEVETLKKAKIEGTQSLRGLDVTIRSEKENLTNELQKEQER-----ISELEI 2509  
QY 235 TEPYFFPAPSKPEMTKADLEQCWDIYFFQFCNFSVSIYESVETITKMLKA-IEDLPK 293

DB 2510 INSFENILQEKQEKVOMKE-----KSSSTAMEMLOQTQKELNE 2548  
QY 294 QDKAPDHGGLISKMLPGQDRGLCGELDQNLNLS---RCFKFHEKCKCQCAHLSBDCPDVPAL 350  
DB 2549 RVAALHN-----DOEACKAKAEQNLSQVECLEL-EKAQLLQG----- 2584  
QY 351 HTEDEAIRLVNVSNOQYQGILOMTRKHLEDPAYLVEKMRGQFGWVSELANQAPETEIF 410  
DB 2585 ---LDEAKNNYIVLQSSVKGLIQ-----EVEDGKQKLEKKDEE---ISRLKNOIQDQEQIV 2634  
QY 411 NSIQVVPRIH-----EGNISKQDETMKTD--LSILPSSNFTLKIPL 450  
DB 2635 SKLSQVEGEHQLWKMEQNLRLNLTVELEOKIQVLOSKNASLODTTLE 2680

RESULT 15  
PCT-US95-16216-1  
; Sequence 1, Application PC/TUS9516216  
; GENERAL INFORMATION:  
; APPLICANT: Yen, Timothy J.  
; TITLE OF INVENTION: Nucleic Acid Encoding a Transiently  
; TITLE OF INVENTION: Expressed Kinetochore Protein, and Methods of Use  
; NUMBER OF SEQUENCES: 4  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Dann, Dorfman, Herrell and Skillman  
; STREET: 1601 Market Street Suite 720  
; CITY: Philadelphia  
; STATE: PA  
; COUNTRY: USA  
; ZIP: 19103-2307  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: PCT/US95/16216  
; FILING DATE:  
; CLASSIFICATION:  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/353,700  
; FILING DATE: 09-DEC-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Reed, Janet E.  
; REGISTRATION NUMBER: 36,252  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (215) 563-4100  
; TELEFAX: (215) 563-4044  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 3248 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: not relevant  
; TOPOLOGY: not relevant  
; MOLECULE TYPE: protein  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
PCT-US95-16216-1

Query Match 5.1%; Score 128.5; DB 5; Length 3248;  
Best Local Similarity 18.8%; Pred. No. 0.012;  
Matches 99; Conservative 85; Mismatches 165; Indels 177; Gaps 22;

QY 27 KUSHCAPTWKDKTAISENLK-----SFSEVGE-----IDADE---EVKKALGTGIQ 69  
DB 2230 KQLHIAEKLERERENDSKDKVENLERELQSEENQELVILDAENSKAEVETLKTQIEE 2289  
QY 70 M-----KIMMERKEKEH-----TNLMSTLKKCKREKQKALKLLNEV 105  
DB 2290 MARSLKIFDLVTLRSEKENLTKQIEKQGLSELDKLLSSPKSLLEEKEQA-----EI 2344





**THIS PAGE BLANK (USPTO)**

---

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:44:34 ; Search time 32.87 Seconds  
(without alignments)  
1394.421 Million cell updates/sec

Title: US-09-722-544A-4

Perfect score: 2529

Sequence: 1 MRTWDSNSGNMKPPLLVI.....FGYVAKALQHFKEHFKTW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : PIR 71:\*

1: pir1:\*

2: pir2:\*

3: pir3:\*

4: pir4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	434	17.2	449	1 A41386	clusterin precurs
2	416	16.4	451	2 I50131	clusterin - quail
3	410.5	16.2	445	2 A40018	clusterin precurs
4	406	16.1	449	2 S07714	T64 protein precurs
5	403	15.9	448	2 A40714	clusterin precurs
6	402	15.9	448	2 I56335	clusterin precurs
7	400.5	15.8	439	2 A35744	apolipoprotein J -
8	400.5	15.8	446	2 A42108	clusterin precurs
9	380	15.0	447	2 A27205	clusterin precurs
10	168	6.6	191	2 I48174	sulfated glycoprot
11	160	6.3	1738	2 T14867	interaptin - slime
12	151.5	6.0	1091	2 T34107	hypothetical prote
13	151	6.0	1526	2 T41522	myosin ii - fissio
14	146	5.8	1156	2 B70356	chromosome assembl
15	145.5	5.8	594	1 A44073	CIK1 protein - yea
16	145	5.7	1388	2 S70633	serine/threonine-s
17	144	5.7	3225	2 I52300	giantin - human
18	144	5.7	3259	1 A56539	giantin - human
19	142.5	5.6	2663	1 S28261	centromere protein
20	141.5	5.6	3433	1 S29381	utrophin - human
21	141	5.6	1005	2 A64465	hypothetical prote
22	140.5	5.6	2166	2 G70163	hypothetical prote
23	139	5.5	1955	2 T30934	myosin-like protei
24	137	5.4	1084	2 T13329	hypothetical prote
25	137	5.4	1300	2 I53799	CGI protein - huma
26	137	5.4	1940	1 S04090	myosin heavy chain
27	136	5.4	1132	2 T00259	hypothetical prote
28	135	5.3	1060	1 A40264	kinesin-related pr
29	134.5	5.3	1957	2 T38077	hypothetical colle

30	134	5.3	1356	2 S32763	kinectin 1 - human
31	133.5	5.3	1313	2 A48467	myosin heavy chain
32	133.5	5.3	1956	2 T16416	hypothetical prote
33	132.5	5.2	852	2 D72230	conserved hypotet
34	132	5.2	1033	2 T42701	hypothetical prote
35	132	5.2	1133	2 T22976	hypothetical prote
36	132	5.2	1388	2 S74245	serine/threonine-s
37	132	5.2	1413	2 T26467	hypothetical prote
38	132	5.2	1992	2 A47297	myosin heavy chain
39	131.5	5.2	1679	2 S48385	hypothetical prote
40	131	5.2	746	2 T21277	hypothetical prote
41	130.5	5.2	1830	1 S19188	myosin-v - chicken
42	130.5	5.2	1940	1 A24922	myosin heavy chain
43	130	5.1	1538	2 T29095	cardiac muscle fac
44	130	5.1	1590	2 B87754	protein C43E11.3 [
45	130	5.1	1961	1 A61231	myosin heavy chain

#### ALIGNMENTS

RESULT 1

A41386  
clusterin precursor [validated] - human  
N:Alternate names: apolipoprotein J; complement cytolysis inhibitor SP-40; complement TRPM-2/clusterin protein  
C:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Homo sapiens (man)  
C:Date: 03-Apr-1992 #sequence\_revision 17-Nov-1995 #text\_change 08-Dec-2000  
C:Accession: S43646; S04662; A41386; A35833; S34056; A53177; A37816; B37816; PL0136; R:Wong, P.; Taillefer, D.; Lakins, J.; Pineault, J.; Chader, G.; Tenniswood, M.  
Eur. J. Biochem. 221, 917-925, 1994  
A:Title: Molecular characterization of human TRPM-2/clusterin, a gene associated with A:Reference number: S43646; MUID:94237156  
A:Accession: S43646  
A:Molecule type: DNA  
A:Residues: 1-449 <WON>  
A:Cross-references: GB:M64722; NID:g339972; PIDN:AA806508.1; PID:g339973 R:Kirsbaum, L.; Sharpe, J.A.; Murphy, B.; d'Apice, A.J.F.; Classon, B.; Hudson, P.; EMBO J. 8, 711-718, 1989  
A:Title: Molecular cloning and characterization of the novel, human complement-associ A:Reference number: S04662; MUID:89251601  
A:Accession: S04662  
A:Molecule type: mRNA  
A:Residues: 1-449 <KIR>  
A:Cross-references: EMBL:X14723; NID:g30250; PIDN:CAA32847.1; PID:g30251  
A:Note: parts of this sequence, including the amino end of the mature protein, were c R:Jenne, D.E.; Tschopp, J.  
Proc. Natl. Acad. Sci. U.S.A. 86, 7123-7127, 1989  
A:Title: Molecular structure and functional characterization of a human complement cy tis fluid.  
A:Reference number: A41386; MUID:89386692  
A:Accession: A41386  
A:Molecule type: mRNA  
A:Residues: 2-449 <JEN>  
A:Cross-references: GB:M25915; NID:g180619; PIDN:AAA35692.1; PID:g180620 R:de Silva, H.V.; Harmony, J.A.K.; Stuart, W.D.; Gil, C.M.; Robbins, J.  
Biochemistry 29, 5380-5389, 1990  
A:Title: Apolipoprotein J: structure and tissue distribution.  
A:Reference number: A35833; MUID:90344779  
A:Accession: A35833  
A:Molecule type: mRNA  
A:Residues: 34-449 <DES>  
A:Cross-references: GB:J02908; NID:g178854; PIDN:AAA51765.1; PID:g178855 R:Ghiso, J.; Matsubara, F.; Koudinov, A.; Choi-Miura, N.H.; Tomita, M.; Wisniewski, T Biochem. J. 293, 27-30, 1993  
A:Title: Research Communication. The cerebrospinal-fluid soluble form of Alzheimer's A:Reference number: S34056; MUID:93319521  
A:Accession: S34056  
A:Molecule type: protein  
A:Residues: 228-240, 'X', 242-246; 23-24, 'X', 26-34, 'X', 36-38, 'X', 40-41 <GHI> R:James, R.W.; Hochstrasser, A.C.; Borghini, I.; Martin, B.; Pometta, D.; Hochstrasse Arterioscler. Thromb. 11, 645-652, 1991



## C:Genetics:

A:Gene: T64

A:Introns: 29/1; 78/3; 135/3; 275/1; 313/1; 388/3; 450/2

C:Superfamily: clusterin

Query Match 16.4%; Score 416; DB 2; Length 451;  
Best Local Similarity 25.7%; Pred. No. 7.7e-19;  
Matches 118; Conservative 93; Mismatches 180; Indels 68; Gaps 15;

QY 45 LKSFSEVGEDADAEVKKALTGIKQIMMERKEKEHTNLMSTLKKCREKQKALKLLNE 104  
Db 27 LKQSAAGSKYIDAENVANGVKQMTLMDKTSKEQAMLTLEETKKKEAVKLALE 86  
QY 105 VQHELEERLCRESLADSGECRSCLENMRYT--TCQPSNWSVKNKTERFRKIYQF 163  
Db 87 KEQLAEKQEVNETHLSLWEECKPCLKHCHCMRVYSKMGCHSGSLVGRQLEEFNRSPPF 146  
QY 164 LFPFHEDNEKDLPISEKLIIE-----DAQLTQMEDVFSQLTVDVNSLFRNSFVFMQMQ 218  
Db 147 SIWNGERIDDLDRQQRERFEDLEERFGLMEDGVEDIFQDSTQLYGPAFFPR---- 202  
QY 219 EFDQTFQSHFISDLDLTPY--FFPAFSKEPMTKADLEQCWDIPN-----FQFLFC 267  
Db 203 -----TPPGGGRFAP-VPPVORVHL-----VPRRLSRELHPFFQHPM 240  
QY 268 NFSVSIYVESVETTKMLKAIEDLPKQDKAPDH--GGLISKMLP-GQDRGLCGELDONLS 324  
Db 241 HGFRHLFQPLFQTMQHLMDG-----GHGAWEHPLGFGFATSNFSTRMVCREIRNSA 294  
QY 325 RCFKFEKCKQKQAHLSDEC-----PDVPAHLTDELDAIRLVNVSNQOYGQILQMTKHLIE 380  
Db 295 GCLMRDECEKREITAVDCSOTDPVQSQRLEQFEDALRLAERFTRRYDDLLSAFQAEML 354  
QY 381 DTAYLVKMRGQFGVSELANQAP-----ETELIENSIVQVPRHIEGNISKQDETMMDT 434  
Db 355 NTSSLLDLNQRQGWRSRLGNLTQNGDGLQVTVTFK---TPNLEDP--SAPADQVTV 408  
QY 435 LSLIPSNFTYKLTPILEESAESSNFIQYVAKALQHFKEH 473  
Db 409 VOLFDSEPLSLVPGDISWDPRFMEITVAEQALQHYKQN 447

## RESULT 3

A40018

clusterin precursor - dog

N:Alternate names: 80k secreted glycoprotein, renal; complement cytolysis inhibitor

C:Species: Canis lupus familiaris (dog)

C:Date: 20-Mar-1992 #sequence\_revision 20-Mar-1992 #text\_change 13-Aug-1999

C:Accession: A40018

R:Hartmann, K.; Rauch, J.; Urban, J.; Parczyk, K.; Diel, P.; Pillarsky, C.; Appel, D.; Ha

J. Biol. Chem. 266, 9924-9931, 1991

A:Title: Molecular cloning of gp 80, a glycoprotein complex secreted by kidney cells in

A:Reference number: A40018; MUID:91236776

A:Accession: A40018

A:Molecule type: mRNA

A:Residues: 1-445 &lt;HAR&gt;

A:Cross-references: GB:M55251; GB:M38757; NID:g163953; PIDN:AAA30846.1; PID:g163954

A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, w

C:Superfamily: clusterin

C:Keywords: apoptosis; disulfide bond; glycoprotein

F:1-22/Domain: signal sequence #status predicted &lt;SIG&gt;

F:23-226/Domain: clusterin beta chain #status experimental &lt;BCH&gt;

F:227-445/Domain: clusterin alpha chain #status experimental &lt;ACH&gt;

F:86,103,145,277,287,350,370/Binding site: carbohydrate (Asn) (covalent) #status predict

Query Match 16.2%; Score 410.5; DB 2; Length 445;

Best Local Similarity 23.9%; Pred. NO. 1.7e-18;

Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 12 MKPPLLVITVCLLWKDSHCAPTWKDKTAISENLKSFSEVGEDAD-----EEVKKALT 65

Db 1 MMKTLTLLVGLLL-----TWNGRVLGDOAVSDTELQENSTEGSKYINKEIKNALK 51  
QY 66 GIKQIMMERKEKEHTNLMSTLKKCREKQKALKLLNEVQHEERLCRESLADSG 125  
Db 52 GVKQIKTLTQOTNEERKSLNLEBAKKKEDALNDTKSETKKASQGVCDNDMMALWE 111  
QY 126 ECRSCLENNCMRYT--TCQPSNWSVKNKTERFRKIYQFLFPFHEDNEKDLPISEKLIIE 184  
Db 112 ECKPCLKQTCMKFYARVCRSGSLVGHQLEEFNLQSSPFVFWMGDR-----IDSLEN 165  
QY 185 DAQLTQMEDVFSQLTVDVNSLFRNSFVFMQOQFQDQTFQSHFIS-----DLDLTPPYFF 240  
Db 166 DROQTHALDV-----MODSNRASSI-----MDELFDQFRFFTRPDQTYHYSP--F 209  
QY 241 PAFSKEPMTKADLEQCWDI-----PNFOLFCSFVSIVSEVSETTKMLKALTE--- 289  
Db 210 SLFQRRPFENPKFRIARNIIPPRFQPLNFDHMFQFFDMIHQA--QOAMDVNLHRIPIYF 268  
QY 290 --DLPKQDKAPDHGGLISKMLPQDRGLCGELDONLSRCFKFHEKCKQKQAHLSDEC--- 344  
Db 269 PIEFPEED-----NRTVCKEIRHNSTGCLMKKQCEKQCEILSVDCSSN 312  
QY 345 -PDVPAHLTDELDAIRLVNVSNQOYGQILQMTKHLIEDTAYLVKMRGQFGVSELANQA 403  
Db 313 NPAQVQLRQELSNLSQIAEKFTKLYDELLOQYQEKMFNTSSLKQLNEQFSWVSQLANLT 372  
QY 404 PETEIIFNISQIV-PRHIEGNISKQDETMMDLSLIPSNFTYKLTPILEESAESSNFIQYV 462  
Db 373 QSEDFFLYQVTVTQSDSNVPGFTKV--VKLFSDSPITVMIPEAVSRNPNRMEIV 430  
QY 463 VAKALQHFKE 472  
Db 431 AEKALQEQYRQ 440

RESULT 4  
S07714  
T64 protein precursor - Japanese quail  
C:Species: Coturnix coturnix japonica (Japanese quail)  
C:Date: 29-Jan-1993 #sequence\_revision 29-Jan-1993 #text\_change 13-Aug-1999  
C:Accession: S07714  
R:Michel, D.; Gillet, G.; Volovitch, M.; Pessac, B.; Calothy, G.; Brun, G.  
Oncogene Res. 4, 127-136, 1989  
A:Title: Expression of a novel gene encoding a 51.5 kD precursor protein is induced b  
A:Reference number: S07714; MUID:89239492  
A:Accession: S07714  
A:Molecule type: mRNA  
A:Residues: 1-449 <MIC>  
A:Cross-references: EMBL:X15825; NID:g62594; PIDN:CAA33823.1; PID:g62595  
C:Superfamily: clusterin  
F:1-18/Domain: signal sequence #status predicted <SIG>  
F:19-449/Product: T64 protein #status predicted <MAT>

Query Match 16.1%; Score 406; DB 2; Length 449;  
Best Local Similarity 25.7%; Pred. No. 3.2e-18;  
Matches 117; Conservative 91; Mismatches 180; Indels 68; Gaps 15;

QY 45 LKSFSEVGEDADAEVKKALTGIKQIMMERKEKEHTNLMSTLKKCREKQKALKLLNE 104  
Db 27 LKQSAAGSKYIDAENVANGVKQMTLMDKTSKEQAMLTLEETKKKEAVKLALE 86  
QY 105 VQHELEERLCRESLADSGECRSCLENMCMRYT--TCQPSNWSVKNKTERFRKIYQF 163  
Db 87 KEQLAEKQEVNETHLSLWEECKPCLKHCHCMRVYSKMGCHSGSLVGRQLEEFNRSPPF 146  
QY 164 LFPFHEDNEKDLPISEKLIIE-----DAQLTQMEDVFSQLTVDVNSLFRNSFVFMQMQ 218  
Db 147 SIWNGERIDDLDRQQRERFEDLEERFGLMEDGVEDIFQDSTQLYGPAFFPR---- 202  
QY 219 EFDQTFQSHFISDLDLTPY--FFPAFSKEPMTKADLEQCWDIPN-----FQFLFC 267

```
Db 203 -----TPPFGGFEAF-VPPVQVRVHL-----VPRRLSLRELHPFFQHPM 240
Qy 268 NFSYIYESVSETITKMLKATEDLPKODKADH--GGLSKMLP-GODRGLCGELDONLS 324
Db 241 HGFRLFPQLEMTQHMLDG-----CHGAWHEPLGGFATERSNFSTRVDRWCRIENSA 294
Qy 325 RCFRFEKQKQCAHLSDEC-----PDVPALHTEDEAIRLVNVSNOQYGOQLQMTKRHLE 380
Db 295 GCLMRDECEKREILAVDCSQTPVQSOLREQFEDALRAERTRRYDDLLSAFAEML 354
Qy 381 DTAVLVEKMRQGFQWSELANQAP-----ETELIFNSIQVVPRIHEGNIKQDETMTWD 434
Db 355 NTSSLLDQLNRFQGWVSRNLGNTQGNDFLQVTTVFESK---TPNLEDP--SAPADTQVT- 408
Qy 435 LSLPSSNFTLKIPLEESAESSNFIQYVAKALQHF 470
Db 409 VQLFDESEPLSLTVFCDISWDDPRFWEIVPEQALQHY 444

RESULT 5
clusterin precursor - mouse
N:Alternate names: sulfated glycoprotein 2
N:Contains: clusterin alpha chain; clusterin beta chain
C:Species: Mus musculus (house mouse)
C:Date: 12-May-1994 #sequence_revision 12-May-1994 #text_change 13-Aug-1999
C:Accession: A40714; JN0699
R:French, L.E.; Chonn, A.; Ducrest, D.; Baumann, B.; Belin, D.; Wohlwend, A.; Kiss, J.Z.
J. Cell Biol. 122, 1119-1130, 1993
A:Title: Murine clusterin: molecular cloning and mRNA localization of a gene associated
A:Reference number: A40714; MUID:93359508
A:Accession: A40714
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:L08235; MID:g192596; PIDN:AAA37422.1; PID:g192597
R:Lee, K.H.; Ji, Y.M.; Lim, H.M.; Lee, S.C.; You, K.H.
Biochem. Biophys. Res. Commun. 194, 1175-1180, 1993
A:Title: Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA from testis of
A:Reference number: JN0699; MUID:93356785
A:Accession: JN0699
A:Molecule type: mRNA
A:Residues: 1-448 <LEP>
A:Note: the authors translated the codon CGC for residue 66 as Ala
C:Comment: This protein is involved in many different cellular processes other than those
C:Superfamily: clusterin
C:Keywords: apoptosis; disulfide bond; glycoprotein; spermatogenesis
F:1-21/Domain: signal sequence #status predicted <SIG>
F:22-226/Domain: clusterin beta chain #status predicted <BCH>
F:227-448/Domain: clusterin alpha chain #status predicted <MAT>
F:227-448/Domain: clusterin alpha chain #status predicted <ACH>
F:102,144,290,327,353,373/Binding site: carbohydrate (Asn) #status predicted

Query Match 15.9%; Score 403; DB 2; Length 448;
Best Local Similarity 23.4%; Pred. No. 5e-18;
Matches 112; Conservative 109; Mismatches 196; Indels 62; Gaps 14;

Qy 16 LILVFIIVCLWLKDSHCAPTWKDKTAISENLKSFSEVEIDA-----DEVKKALTGIKQ 69
Db 3 ILLLCVALLI-----WDNGMVLGEQVSDNELQELSTQGSRYINKEIQNAVQGVKH 54

Qy 70 MKIMMERKEHTNLMSTLKCKREEKQEAALKLLNEVQEHLEBEERLCRESLADSGEGRS 129
Db 55 IKTLIEKTNAERKSLNLSLEAKKKEDALEDTROSEMKLKAPPEVCNMTMALWEECKP 114

Qy 130 CLENNCMRIYT--TCQPSWSVSNKIKRFRFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQL 188
Db 115 CLKHTCMKFIYARVCRSGSLVGQOLEEFLNQSSPFFYFWNGDR-----IDSLESDRQ 168

Qy 189 TO-----MEDVFSQLTVDVNSLNRSPVFRMQQEQDFQTSFHSIDTDLTPYFPF--- 241
Db 169 SQVLDAMQDSFARASGIITLFDRE--FAR-----ELHDPHYFS-----PIGPPH 214
```

```
Qy 242 ---AFSKPEMTKADLEOC-WDIPNFFQPCNFVSIVSEYSETITKMLKAIEDLPKODKA 297
Db 215 PHFLYPKSRLVRSILMSPSHYGPSPFSEHMFQFFEMIHQAQAMDMVQLHSPAFQPDVD-- 272
Qy 298 PDHGGLLSKMLPGQ-DRGLCGELDONLSRCFKFHEKQKQCAHLSDEC-----PDVPALHT 352
Db 273 -----FLREGEDDRTVCKEIRNRSTGCLKMKGQCEKQCEILSVDCSTNNPAQANLRQ 324
Qy 353 ELDEAIRLVNVSNOQYGOQLQMTKRKHLEDYALVEKMRGQFGWVSELANQAPETELIFNS 412
Db 325 ELNDSLQVAERLTEQYKELLQSFQSKMLNTSSLEQLNDQFNWVSQANLTQGEDKYLR 384
Qy 413 IQVVPRIHEGNIKQDETMTDLISLPSNFTLKIPLEESAESSNFIQYVAKALQHF 471
Db 385 VSTV-TTHSSDSEVPSPRVTEVYVVKLFDSDPITVVLPEEVSKDNPKFMDTVAEKALQEYR 442

RESULT 6
apolipoprotein J - mouse
C:Species: Mus sp. (mouse)
C:Date: 26-Jul-1996 #sequence_revision 26-Jul-1996 #text_change 13-Aug-1999
C:Accession: I56335
R:Jordan-Starck, T.C.; Lund, S.D.; Witte, D.P.; Atonow, B.J.; Ley, C.A.; Stuart, W.D.
J. Lipid Res. 35, 194-210, 1994
A:Title: Mouse apolipoprotein J: characterization of a gene implicated in atherosclerosis
A:Reference number: I56335; MUID:94223204
A:Accession: I56335
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-448 <RES>
A:Cross-references: GB:S70244; MID:g546504; PIDN:AAB30623.1; PID:g546505
C:Genetics:
A:Gene: ApoJ
C:Superfamily: clusterin

Query Match 15.9%; Score 402; DB 2; Length 448;
Best Local Similarity 24.1%; Pred. No. 5.8e-18;
Matches 115; Conservative 108; Mismatches 201; Indels 54; Gaps 15;

Qy 12 MKPPLLFIIVCLWLKDSHCAPTWKDKTAISENLKSFSEVEIDADEVKALTGIKQ 70
Db 1 MKILLCVALLI--DNGMV---LGEQVSDNELQELSTQGSRYINKEIQNAVQGVKH 55

Qy 71 KIMMERKEHTNLMSTLKCKREEKQEAALKLLNEVQEHLEBEERLCRESLADSGEGRS 130
Db 56 KTLIEKTNAERKSLNLSLEAKKKEDALEDTROSEMKLKAPPEVCNMTMALWEECKP 115

Qy 131 LENNCMRIYT--TCQPSWSVSNKIKRFRFRKIYQFLFPFHEDNEKDLPISEKLIIEADAQL 189
Db 116 LKHTCMKFIYARVCRSGSLVGQOLEEFLNQSSPFFYFWNGDR-----IDSLESDRQ 169

Qy 190 Q-----MEDVFSQLTVDVNSLNRSPVFRMQQEQDFQTSFHSIDTDLTPYFPF--- 241
Db 170 QVLDAMQDSFARASGIITLFDRE--FAR-----ELHDPHYFS-----PIGPPH 215

Qy 242 ---AFSKPEMTKADLEOC-WDIPNFFQPCNFVSIVSEYSETITKMLKAIEDLPKQKAP 298
Db 216 HFLYPKSRLVRSILMSPSHYGPSPFSEHMFQFFEMIHQAQAMDMVQLHSPAFQPDVD-- 272

Qy 299 DHGGLSKMLPGQ-DRGLCGELDONLSRCFKFHEKQKQCAHLSDEC-----PDVPALHTE 353
Db 273 -----FLREGEDDRTVCKEIRNRSTGCLKMKGQCEKQCEILSVDCSTNNPAQANLRQ 325

Qy 354 LDEAIRLVNVSNOQYGOQLQMTKRKHLEDYALVEKMRGQFGWVSELANQAPETELIFNSI 413
Db 326 LNDLSQVAERLTEQYKELLQSFQSKMLNTSSLEQLNDQFNWVSQANLTQGEDKYLR 385

Qy 414 QVVPRIHEGNIKQDETMTDLISLPSNFTLKIPLEESAESSNFIQYVAKALQHF 471
Db 386 STV-TTHSSDSEVPSPRVTEVYVVKLFDSDPITVVLPEEVSKDNPKFMDTVAEKALQEYR 442
```

## RESULT 7

A35744  
clusterin precursor - bovine  
N:Alternate names: complement cytolysis inhibitor; glycoprotein III  
N:Contains: clusterin alpha chain; clusterin beta chain  
C:Species: Bos primigenius taurus (cattle)  
C>Date: 05-Oct-1990 #sequence\_revision 05-Oct-1990 #text\_change 13-Aug-1999  
C:Accession: A35744  
R:Palmer, D.J.; Christie, D.L.  
J. Biol. Chem. 265, 6617-6623, 1990  
A:Title: The primary structure of glycoprotein III from bovine adrenal medullary chromaffin granules  
A:Reference number: A35744; MUID:90216681  
A:Accession: A35744  
A:Molecule type: mRNA  
A:Residues: 1-439 <PAL>  
A:Cross-references: GB:J05391; NID:gl63114; PIDN:AAA30554.1; PID:gl63115  
A:Note: parts of this sequence, including the amino ends of the alpha and beta chains, were deleted in the original publication  
C:Superfamily: clusterin  
C:Keywords: apoptosis; disulfide bond; extracellular protein; glycoprotein  
F:1-19/Domain: signal sequence #status predicted <SIG>  
F:20-221/Domain: clusterin beta chain #status experimental <BCH>  
F:20-221,222-439/Product: clusterin #status experimental <MAT>  
F:222-439/Domain: clusterin alpha chain #status experimental <ACH>  
F:80,97,139,283,320,346,366/Binding site: carbohydrate (Asn) (covalent) #status predicted

## Query Match

Best Local Similarity 15.8%; Score 400.5; DB 2; Length 439;  
Matches 114; Conservative 97; Mismatches 183; Indels 71; Gaps 15;

QY 34 TWKDKTAISE-NLKSFSEVEIDADEVKKALTGIKIMMERKEKEHTNLMSTLKCR 92  
DB 13 SWEGSAISDEKQEMSTEGSKYVNEIKNALKEVQIKTQIBQTEERKLLLSLEAK 72  
QY 93 EEKQKALLNVEQHELEERLCRESLADSGECRSCLNENCMRIYT-TCQPSWSVKN 151  
DB 73 KKEDALNDRSENKLLASQGCNENMTALMEECKPCLKQTCMKFYARVCRSGSLVGH 132  
QY 152 KIRFRPKYQVFFPHEDNEKDLPISEKLEDAQLTQMEDVFSQLTVDVNSLFRSN 211  
DB 133 QLEEFNLQSSPFFWINGDR-----IDSLMENDREQSHVMDV-----MEDSFTRAS 179  
QY 212 VFQMQQEDQTFQSHFI---SDTLTEPY-----FF---PAFSKEPTKADLEQC 256  
DB 180 I-----MDELQDFRFLRRPDQTOYYSFSGSFRGSLFFNPKSRFARNVMPFLLPEF 232  
QY 257 WDTIPNFFLFCNFSVIYESVSETITPKMLKAIEDLPKQKADPHGGLISKMLPGQDRLC 316  
DB 233 ----NPHDVQFP-----YDWHQAQQAQDAHLQ-----RTPYHPTM-EFTENDRTVC 277  
QY 317 GELQNLKRCFHEKCKQCAHLSEDC-----PDVPAHLTEDEAIRLVNSNQOYQIL 372  
DB 278 KEIRHNSTGLRMKQCEKCEKILEYDCSASNPQTQTLRQQLNASLQAEKFSRLYDQLL 337  
QY 373 QMTRKHELDATYLVKMRGQFGWVSELANQAPETE-----IIFNSIQVVPRIHGNISK 426  
DB 338 QSVQKMLNTSALLKQNLNQFTWVSQLANLTQSDQHYLOVFTVNSHNSDPSIPSG----- 393  
QY 427 QDETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQHF 471  
DB 394 ----LTKVIVKLFNSFPITVVPQVSSPNFMENVAEKALQOYR 433

## RESULT 8

A42108  
clusterin precursor - pig  
N:Alternate names: complement cytolysis inhibitor; CP40 protein  
C:Species: Sus scrofa domestica (domestic pig)  
C>Date: 04-Mar-1993 #sequence\_revision 18-Nov-1994 #text\_change 13-Aug-1999  
C:Accession: A42108; JC5535; PC4475  
R:Diemer, V.; Hoyle, M.; Baglioni, C.; Millis, A.J.  
J. Biol. Chem. 267, 5257-5264, 1992

A:Title: Expression of porcine complement cytolysis inhibitor mRNA in cultured aortic  
A:Reference number: A42108; MUID:92184774  
A:Accession: A42108  
A:Molecule type: mRNA  
A:Residues: 1-446 <DIE>  
A:Cross-references: GB:M84639; NID:gl64408; PIDN:AAA31013.1; PID:gl64409  
A:Experimental source: aortic smooth muscle cells  
A:Note: sequence extracted from NCBI backbone (NCBIN:87354, NCBI:87356)  
R:Ogawa, S.; Ishibashi, Y.; Sakamoto, Y.; Kitamura, K.; Kubo, M.; Sakai, T.; Inoue, K.  
Biochem. Biophys. Res. Commun. 234, 712-718, 1997  
A:Title: The glycoproteins that occur in the colloids of senescent porcine pituitary  
A:Reference number: JC5535; MUID:97318844  
A:Accession: JC5535  
A:Molecule type: DNA  
A:Residues: 1-446 <OGA>  
A:Accession: PC4475  
A:Molecule type: protein  
A:Residues: 58-66;68-77;229-247;249-251;408-436 <OG2>  
A:Experimental source: pituitary gland  
C:Superfamily: clusterin  
F:1-22/Domain: signal sequence #status predicted <SIG>  
F:23-227/Domain: clusterin beta chain #status predicted <BCH>  
F:23-227,228-446/Product: clusterin #status predicted <MAT>  
F:228-446/Domain: clusterin alpha chain #status predicted <ACH>

## Query Match

Best Local Similarity 15.8%; Score 400.5; DB 2; Length 446;  
Matches 118; Conservative 98; Mismatches 196; Indels 71; Gaps 16;

QY 16 LLVFIIVCLLWLKDSHCAPTWK-----DKTAISENLKSFSEVEIDADEVKKALTG 58  
DB 4 LLLLVGLLL-----TWGNPVGDKALSDLEQEMSTEGSKYVNEIKNALKEVK 54  
QY 69 QMKIMMERKEHTNLMSTLKCRKQKALKEVQHELEERLCRESLADSGECR 128  
DB 55 QIKTLIEQSSNEERKLLSLEAKKKEDALNDRTRTETKLAGSGLCNETMMALWEECK 114  
QY 129 SCLNENCMRIYT-TCQPSWSVKNKIERFRKIYQVFFPHEDNEKDLPISEKLEEDAQ 187  
DB 115 PCLKQTCMKFYARVCRSGSLVGHQLEEFNLQSSPFFWINGDR-----IDSLMENDRQ 168  
QY 188 LTQMEDVFSQLTVDVNSLFRSNFVROMQEQDFDTQTFQSHFIS----DTDLTEPY----- 238  
DB 169 QSHVMDI-----MEDSFNRSNI-----MDELQDFRFRNRPFFDTQTFSPGSSHR 214  
QY 239 ----FFPAFSKEPTKADLEQCWDIPNFFLFCNFSVIYESVSETITPKMLKAIEDLPKQD 295  
DB 215 GSLFFNPKSRFARNIMPFFLEFIDL-NYHDMFQFP-----FDMIHQQAQMDAHLRIYH- 268  
QY 296 KAPDHGGLISKMLP--GODRGLCGELDONLSRCFKFHEKCKQCAHLSEDC-----PDVPA 349  
DB 269 -FPEAG-----VPNSNDRAVCKEIRHNSTGLRMKQCEKCREILSVDCSASNSQ 321  
QY 350 LHTEDEAIRLVNSNQOYQILQMTKRKHELDATYLVKMRGQFGWVSELANQAPETEI 409  
DB 322 LRQELYTSLOMAEKFSKLYDQLQSQYQKMLNTSLLKQNLNQFTWVSQLANLTQDNDRY 381  
QY 410 ENSIQVVPRIHGNISKQDETMMTDLISLPSSNFTLKIPLESAESSNFYGVVAKALQ 469  
DB 382 YLOVTVV-NSHGSDPSVPVSGLTQVVKVLFDSYPTLTIPIQEVN--DPKFMETVAEALQ 438  
QY 470 FKE 472  
DB 439 YRQ 441

## RESULT 9

A27205  
clusterin precursor - rat  
N:Alternate names: SCP-2; sulfated glycoprotein 2; TRPM-2  
C:Species: Rattus norvegicus (Norway rat)  
C>Date: 21-May-1988 #sequence\_revision 21-May-1988 #text\_change 13-Aug-1999

Db 215 PHFLYPKSLRSLRSLMPLSHVGPLSFHNMFPDPFDMTLHQAAQMDVOLHSPALQFDPDVFEL 277

QY 298 PDHGGLISKMLPCQ-DRGLCGELDONLSRCFRFHEKCKCQKQAHLSDC-----PDVPALHT 352

Db 275 KE-----GEDPTVCKEIRHNSGTCLKMKGCQCKQELLSDVDCSTNNPQAQANLRQ 324

QY 353 ELDEAIRLVNVSNOQYQGQIQLQMKRHELDYAILVEKMRQFGVWSYELAN--QAPETETIIF 410

Db 325 ELNDSLOVAERLTQQYNELLHSLQSKMLNTSSLEQLNDOQFTWVSQANLITQGDQYLRV 384

QY 411 NSTQVPRPRIHEGNISKODEPMTWDTSLTPSSNTTLKIPLEESAESNFTGYVYAKALQHF 470

Db 385 STVTT-----HSSDSEVPSTVTVVWKLFDSDPTTVVLPPEVSKDNPKFMDTVAEKALQEQY 440

QY 471 K 471

Db 441 R 441

RESULT 10

I48174

sulfated glycoprotein 2 - golden hamster (fragment)

C:Species: Mesocricetus auratus (golden hamster)

C:Date: 27-Feb-1997 #sequence\_revision 27-Feb-1997 #text\_change 13-Aug-1999

C:Accession: I48174

R:Duguid, J.R.; Bohmont, C.W.; Liu, N.G.; Tourtellotte, W.W.

Proc. Natl. Acad. Sci. U.S.A. 86, 7260-7264, 1989

A:Title: Changes in brain gene expression shared by scrapie and Alzheimer

A:Reference number: I48174; MUID:89386721

A:Accession: I48174

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: mRNA

A:Residues: 1-191 <RES>

A:Cross-references: GB:M26640; NID:g191450; PIDN:AAA37102.1; PID:g191451

C:Superfamily: clusterin

Query Match 6.6%; Score 168; DB 2; Length 191;

Best Local Similarity 35.1%; Pred. No. 0.001;

Matches 33; Conservative 22; Mismatches 35; Indels 4; Gaps

QY 312 DRGLCGELDONLSRCFRFHEKCKCQKQAHLSDC-----PDVPALHTELDEAIRLVNVSNOQ 367

Db 69 DRVCKEIRHNSGTCLKMKGCQCKQELLSDVDCSANNPQAQHLRQELNDSLOVAERLTQR 120

QY 368 YGQIQLQMKRHELDYAILVEKMRQFGVWSYELAN 401

Db 129 YNELLHSLQTKMLNTSSLEQLNEQFNWVSQAN 162

RESULT 11

T14867

Interactin - slime mold (Dictyostelium discoideum)

C:Species: Dictyostelium discoideum

C:Date: 20-Sep-1999 #sequence\_revision 20-Sep-1999 #text\_change 11-May-2000

C:Accession: T14867

R:Rivero, F.J.; Kuspa, A.; Brokamp, R.; Matzner, M.; Noegel, A.A.

J. Cell Biol. 142, 735-750, 1998

A:Title: Interactin, an actin-binding protein of the alpha-actinin superfamily.

A:Reference number: T18248; MUID:98365468

A:Accession: T14867

A:Status: preliminary; translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-1738 <RIV>

A:Cross-references: EMBL:AF057019; NID:g3549260; PID:g3549261; PIDN:AAC3494

C:Genetics:

A:Gene: abpD

A:Introns: 173/2; 1680/1

Query Match 6.3%; Score 160; DB 2; Length 1738;

Best Local Similarity 20.9%; Pred. No. 0.049;

Query Match 6.3%; Score 160; DB 2; Length 1738;  
Best Local Similarity 20.9%; Pred. No. 0.049;



Matches 102; Conservative 88; Mismatches 171; Indels 128; Gaps 21;

QY 36 KDKTAISENL-----KSFSEVGEIDADEEVKALTGKQIMMER----- 76  
Db 667 KDNQTIMEQNLKQSEKDEIKLSNQEOQDEKINNNLLEKEKDKCLIERINQOULLEN 726  
QY 77 -----KEKHTNLMSLTKKREEKQKALKLLNEVQHEEER 114  
Db 727 IDLSKYQOULLLEPFENKLNSSKE-----NQLNELQSKODERFNQDND--EKLEKEKQ 779  
QY 115 LCRESLADSWGECRSLNENCMRIYTCQPSWSVKNKIERFRKIIYQVFLFPFHEDNEKD 174  
Db 780 L--QSIDEFNOYK-----QOQLSSNSN-IDQQLQSTIISELSEKQEKELN 822  
QY 175 LPISKELEDAOLQTOQMEDVFSOLT-----VDVNSLFNRSFNVRQMOQOEDQTFOSH 227  
Db 823 ---DSKLEKEKQLOQOQFQDQLNKQKDHQDQLELEKQ---UKQLOQOEDQDQNETN 876  
QY 228 FISDTLTPYFPFAPSKPMTKADLE-----QCWDIPNFFQFCNFSVIVESV 277  
Db 877 QSIENQNLQOQNL---INKENLNEKEQELLKQNLQNOQIEKIQFDQOEFKQNSINIELV 933  
QY 278 SEITTKMLKAIED---LPKQDKA---PDHGLISKMLPGQDGLGELQDNLNLSRCFKFHE 331  
Db 934 NENKELIQLOQDYDQLKQOQNRNDEKNDLIEK--ENOLKSIQNELNOLIEKNESDHK 991  
QY 332 KCQKCAHLSEDCPD---VPALHTELDE---AIRLVNSNOYGOIL-----Q 373  
Db 992 EQQLKQOSIENDLIERENQIQLOQSLNEQROOQSLSEKQDQNLQIEKNQDFQKEQ 1051  
QY 374 MTRKHLEDYAY-----LVEKMRGQFG--WVSELANOQAPETEIFNSIQVVPRIHEGNSKQD 428  
Db 1052 LKQOSIENDLFEKENQIQLOQSLNEQROOQSLSEKQDQNLQIEKNESDHK 1105  
QY 429 ETNMTDLSI 437  
Db 1106 EQQLKQOSI 1114

RESULT 12  
T34107  
hypothetical protein C18C4.5 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C:Date: 29-Oct-1999 #sequence\_revision 29-Oct-1999 #text\_change 29-Oct-1999  
C:Accession: T34107  
R:Gattung, S.  
submitted to the EMBL Data Library, April 1996  
A:Description: The sequence of C. elegans cosmid C18C4.  
A:Reference number: 221478  
A:Accession: T34107  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-1091 <GAT>  
A:Cross-references: EMBL:U55369; PIDN:RAC25825.1; GSPDB:GN00023; CESP:C18C4.5  
A:Experimental source: strain Bristol N2; clone C18C4  
C:Genetics:  
A:Gene: CESP:C18C4.5  
A:Map position: 5  
A:Introns: 38/3; 95/3; 179/2; 259/3; 301/3; 419/2; 573/3; 613/1; 875/3; 920/2; 959/2; 98

Query Match 6.0%; Score 151.5; DB 2; Length 1091;  
Best Local Similarity 20.5%; Pred. No. 0.095;  
Matches 105; Conservative 99; Mismatches 184; Indels 125; Gaps 25;

QY 36 KDKTAISENLKSFSEVGEID-----ADEEVKALTGKQIMK---IMMERKEKEHTNLM 85  
Db 283 KDNNAI-LNVQLEKDGKIDRQVDLLAESRAQQAEDVDMKERIITSKDDSDNNLL 341  
QY 86 -STLKKCRKEQKALKLLNEVQHEEERLCRESLADSWGECRSLNENCMR-----IV 139  
Db 342 QDELRRTEEKYQQAOKKIENLDETIQEQTOIRD-LGRSLDEAKROLQKMSQORQEEVA 400

QY 140 TTCQPSHSSVKNKIERFRKIIYQVFLFPFHEDNEKDLPISEKLEEDAQLTQ----- 190  
Db 401 ROGEDSARSMEEKATKEEIKKLKSOVLOQOQLEQDLQKRVQE---LTERQKRVLESKA 457  
QY 191 -MEDVFSQLVNLSFNRSFNVRQMOQEF-----DQTFQ----- 225  
Db 458 SVADEGTGLMSSLSNL--REEN--RQYEETRSLOTNIRTLQDEVYQHDQAIWENKNRAE 513  
QY 226 -----SHFIS-----DTLTPYFPFAPSKPMTKADLEQOCWDIPNFFQFCNFS 270  
Db 514 KAEYIEKENHRVQNASSSHADITRLENEKTQMEEALEKADQEKQAIRE-----ASES 568  
QY 271 VSIVE-----SVSTITKMLKAIEDLPKQDKAPDHGGLISKMLPGQDGLGELQDNLNLSRC 326  
Db 569 VRVKKREMTASITSQIQSLKE--KVDS-----LTRELESSRRRM-EQLQEDQTKF 618  
QY 327 FKPERCKCQCOAHLSEDCPDVPALHTELDEAIRLVNSNOYGOILQMTKRLHEDYAYLV 386  
Db 619 LGSHD---ETKEMMKD-----LHEAQDIEKLTNQAGOLKSKNETLTTE-LEDSONLC 668  
QY 387 EKMRGQFGV-----SELANOQAPETEIFNSIQVVPRIHEGNSKQDETMTDLSILPSSN 442  
Db 669 ERLKAQYKADKYEETKVQLREAEADLRLQAA-QILSGNVESEKFSDMOKE----- 719  
QY 443 FTLKIPLEESAESSNFGYVVAKALQHFKEHF 475  
Db 720 --SKIEMERILDNHN-----KELEKLEELK 743

RESULT 13  
T41522  
myosin ii - fission yeast (Schizosaccharomyces pombe)  
C:Species: Schizosaccharomyces pombe  
C:Date: 03-Dec-1999 #sequence\_revision 03-Dec-1999 #text\_change 08-Sep-2000  
C:Accession: T41522  
R:Wood, V.; Rajandream, M.A.; Barrell, B.G.; Rieger, M.  
submitted to the EMBL Data Library, March 1999  
A:Reference number: 222000  
A:Accession: T41522  
A:Status: preliminary; translated from GB/EMBL/DBBJ  
A:Molecule type: DNA  
A:Residues: 1-1526 <WOO>  
A:Cross-references: EMBL:AL049498; PIDN:CAB39901.1; GSPDB:GN00068; SPDB:SPCC645.05C  
A:Experimental source: strain 972h-; cosmid c645  
C:Genetics:  
A:Gene: SPDB:SPCC645.05C  
A:Map position: 3  
C:Superfamily: myosin heavy chain; myosin motor domain homology  
F:80-743/Domain: myosin motor domain homology <WMO>

Query Match 6.0%; Score 151; DB 2; Length 1526;  
Best Local Similarity 20.7%; Pred. No. 0.15;  
Matches 93; Conservative 81; Mismatches 221; Indels 54; Gaps 15;

QY 47 SFSEVGEIDADEEVKALTGKQIMMERKEKEHNLMSLTKKREEKQKALKLLNEVQ 106  
Db 895 SFSETKQ--QENLNQRESASLKQINNELSELEKTSKVETL-----LSEQNELK 942  
QY 107 EHLEEEERLCRESLADSWGECRSLNENCMR-----MRIYTCQPSWSVKNK---IERFF 157  
Db 943 EKLSLEEK-----DLLDTKGELESRLNENATVLSEKAEEFNQCKSLQETIVTKDAELDKLT 998  
QY 158 RKIQVQLFFPHEDNEKDLPISEKLEEDAQLTQMEDVFSQLVNLSFNRSFNVRQMOQ 217  
Db 999 KYISDYTKTIQEMRLTNQKNMEKSIQOQESLSKLRVKKLERE-NSTLISDVSLKQOK 1057  
QY 218 QBFD--QTFQSHFISDTLTPYFPFAPSKPMTKADLEQOCWDIPNFFQFCNFSVIVE 275  
Db 1058 EELSVLKGVOELTINNLEKVNYLEADVQLPKLKELESNDKQOLYQATKKELEA 1117  
QY 276 SVSETITKMLKAIEDLP-KQDKAPDHGGLISKMLPGQDGLGELQDNL-----SRCFKFH 330

Db 1118 KVCECLNNIKSLTELENKEKQCNLSASLYIELQ-----EIHENLLLVSDLENYK 1171

Qy 331 EKQCQCAHLSDCPDVPALHTLDEAIRLVNSNQ-----QYGOILQWTRKHLDEPTA 383

Db 1172 KKEGLQLDL-EGKLDVNTFQELSKKHDLTFNHSLLRQSASVKEKLSLASSENKDL 1230

Qy 384 YVLEKMRQFGWVSLANQAPETELIFNSIQVVPRIHE-GNISKODEMTMTDLSTLPSSN 442

Db 1231 NKVSSLTKOVNELSPKASKVPELE-----RKITNLMHYSQGLKTFEDEKRAKALIASRDN 1285

Qy 443 FTLKPLEESAESAFNIGYVAKALQHEK 471

Db 1286 BELR-SLASELESKRKLEVEYQKVLVEVK 1313

RESULT 14

B70356

Chromosome assembly protein homolog - Aquifex aeolicus

C:Species: Aquifex aeolicus

C:Date: 08-May-1998 #sequence\_revision 08-May-1998 #text\_change 02-Jun-2000

C:Accession: B70356

R:Deckert, G.; Warren, P.V.; Gaasterland, T.; Young, W.G.; Lenox, A.L.; Graham, D.E.; Ov

V.

Nature 392, 353-358, 1998

A:Title: The complete genome of the hyperthermophilic bacterium Aquifex aeolicus.

A:Reference number: A70300; MUID:98196666

A:Accession: B70356

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1156 <AQF>

A:Cross-references: GB:AF000699; NID:g2983238; PIDN:AAC06839.1; PID:g2983243; GB:AE00065

A:Experimental source: strain VFS

C:Genetics:

A:Gene: xcpC

C:Superfamily: chromosome segregation protein SMC1

Query Match 5.8%; Score 146; DB 2; Length 1156;

Best Local Similarity 20.0%; Pred. No. 0.23;

Matches 89; Conservative 89; Mismatches 162; Indels 104; Gaps 17;

Qy 45 LKSFSEVGEID-----ADBEVKKALTGIKQKIMME-----RKEKEHTNLMSTLKK 90

Db 163 IBEISGIGEYERKEKALEELAEVLKIKEIDLILEISNQLKRLKEKEKLEKELQ 222

Qy 91 CREEKQKALKNEVQHELEBEERLCRE--SLADSWGECRSCLENNCMRIYTCQPSWSS 148

Db 223 IKRE-TEAKILLKEKELKLERILNELSLRESLEDTTFOIQENEKEL----- 271

Qy 149 VNKKTIERFRKIYQFLFPFHEDNEK--DLPTSEKLI-BEDAQLTQMEDVFSQLTVDVNS 204

Db 272 --NERERLLKEVNEKIMPKEKVGFTABEIAERSAIKEKELAESENVRVKNLEELINN 329

Qy 205 LENRNFNFVFRQ---MQQEPDQTFQSHFTSDTLTPY-FFPAFSKEPMTKADLEQCWDIP 260

Db 330 LLSDXENLEREVGTQLQLELEK-----LKEEYKSLKEVEREKLRELEEE- 373

Qy 261 NFFQQLFCNF-SVSIYESVSETTKMKLKAIEDLPKQKAPDHGGLISKMLPGQDRGLGGBL 319

Db 374 --ERLKITFDVYKLEEEKEKELTEKLSL-NKEKQLEITQIRANLNKI-----ERI 421

Qy 320 DONLSRCFKEKCKQCAHLSDCPDVPALHTLDEAIRLVNSNQYGOILQWTRKHL 379

Db 422 KEDINKLSEREKIKEKEKEQETKRLKAIKKKEEELRNLTQELNIYEKRLSEVRK 481

Qy 380 E-----DTAYLVKMRQFGQVWSELAN-QAPE----- 405

Db 482 EEVLKEKAIEREVSFSFSDVDFDKIGVYGSVELIRVKNPEHITATEVAGGRLKFI 541

Qy 406 ----TEIFNSIQVVPRIHEGNIS 425

Db 542 VVEDEVAKECICQLAKRMNLGRFS 565

RESULT 15

A44073

CIK1 protein - yeast (Saccharomyces cerevisiae)

N:Alternate names: protein YM9646.11; protein YMR198w

C:Species: Saccharomyces cerevisiae

C:Date: 30-Sep-1993 #sequence\_revision 08-Mar-1996 #text\_change 21-Jul-2000

C:Accession: A44073; S50927

R:Page, B.D.; Snyder, M.

Genes Dev. 6, 1414-1429, 1992

A:Title: CIK1: a developmentally regulated spindle pole body-associated protein impor

A:Reference number: A44073; MUID:92354905

A:Accession: A44073

A:Molecule type: DNA

A:Residues: 1-594 <PAG>

A:Cross-references: EMBL:M96439; NID:g171221; PIDN:AAA34494.1; PID:g171222

A:Note: sequence extracted from NCBI backbone (NCBI:110670, NCBI:P:110672)

R:Pearson, D.; Bowman, S.

submitted to the EMBL Data Library, January 1995

A:Reference number: S50917

A:Accession: S50927

A:Molecule type: DNA

A:Residues: 1-594 <PEA>

A:Cross-references: EMBL:Z47815; NID:g642280; PIDN:CAA87820.1; PID:g642291; GSPDB:GNO

C:Genetics:

A:Gene: SGD:CIK1; MIPS:YMR198w

A:Cross-references: SGD:S0004811; MIPS:YMR198w

A:Map position: 13R

C:Superfamily: CIK1 protein

C:Keywords: coiled coil

Query Match 5.8%; Score 145.5; DB 1; Length 594;

Best Local Similarity 20.9%; Pred. No. 0.11;

Matches 84; Conservative 71; Mismatches 138; Indels 109; Gaps 17;

Qy 26 LKDSHCAPTWDKTAISENLKSFSEVG---EIDADEEVKKA----- 63

Db 147 LKDNQCDLQRNKVELSSRNIVSMHAKVQVEFENDLEELSNKAKRWTYKLMVEVNLKPDE 206

Qy 64 -LTG-IKQKIMME---RK-----EKEHTNLMSTLKKCKREEKQEAL-KLLNEVQHELE-EE 112

Db 207 LUTDEMRLQKTEFEVBNRKLFTLQNEENECNKKELDKKFEIKFKVKNDAEIELDGEQ 266

Qy 113 ERLCR-ESLADSWGECRSCLENNCMRIYTCQPSWSSVKNKIERFFRKIYQFLFPFHED 170

Db 267 ERLSKVLKDLQDTHGELKENIK-----TC-----RDEFNDFEKKRIGAEVNFHSM 311

Qy 171 NEKDLPISKELTEDAQLTQMEDVFSQLTVDVNSLNFNSFNVRQMQQEFQDTFQSHFIS 230

Db 312 ELAVVPLKLLASTQALTVQEEKKQVEGEANNKKYVNELEKVVQDEL----- 361

Qy 231 PPDLPTEPVFFPAFSKEPMTKADLE-----QCWDIPNFTOLFQCNFVS-----IYESVSET 280

Db 362 -----YTRQNLATSTEEIKGYTRCFAYANEROMPDEFHINYVDRCICENSGE- 408

Qy 281 ITKMLKAIEDLPKQKAPDHGGLISKMLPGQDRGLGGBLQNLSCFCFKEKCKQCAHL 340

Db 409 --KRVQVDFRVVLEETHKHDKRLYNCEIPFLE-----KYISKLNCSIIIV 451

Qy 341 SEDCDPVPALHTLDEAIRLVNSNQYGOILQWTRK--HLE 380

Db 452 VSQQTPAPMKKTLQLLI-----EQYGENYKMTLNLILHLD 486

Search completed: July 2, 2002, 11:44:36

Job time: 260 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:50:18 ; Search time 17.51 Seconds  
(without alignments)  
1054.782 Million cell updates/sec

Title: US-09-722-544A-4

Perfect score: 2529

Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHFKEFKTW 477

Scoring table:

BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	434	17.2	449	1	CLUS_HUMAN	P10909 homo sapien
2	416	16.4	451	1	CLUS_COTJA	P14018 coturix co
3	413.5	16.4	447	1	CLUS_RABIT	O9xsc5 oryctolagus
4	410.5	16.2	445	1	CLUS_CANFA	P23473 canis famli
5	403	15.9	448	1	CLUS_MOUSE	O06890 mus musculus
6	400.5	15.8	439	1	CLUS_BOVIN	P17697 bos taurus
7	400.5	15.8	446	1	CLUS_PIG	Q29549 sus scrofa
8	380	15.0	447	1	CLUS_RAT	P05371 rattus norv
9	168	6.6	191	1	CLUS_MESAU	P14683 mesocricetu
10	151	6.0	1526	1	MY52_SCHPO	Q9us16 schizosacch
11	145.5	5.8	594	1	C1K1_YEAST	Q01649 saccharomyc
12	144.5	5.7	1938	1	MYHD_HUMAN	Q9ukx3 homo sapien
13	142.5	5.6	2663	1	CENE_HUMAN	Q02224 homo sapien
14	141.5	5.6	3433	1	UTRO_HUMAN	P46939 homo sapien
15	141	5.6	1005	1	RA50_METJA	O58718 methanococc
16	137	5.4	1940	1	MYH3_HUMAN	P11055 homo sapien
17	136	5.4	2230	1	G0G4_HUMAN	Q13439 homo sapien
18	135	5.3	1060	1	E851_XENLA	P28025 xenopus lae
19	134.5	5.3	1957	1	O86_SCHPO	Q10411 schizosacch
20	134	5.3	5430	1	ACF7_HUMAN	O9upn3 homo sapien
21	132.5	5.2	852	1	RA50_THEMA	Q9xlx1 thermotoga
22	131.5	5.2	1679	1	Y109_YEAST	P40457 saccharomyc
23	130.5	5.2	1940	1	MYH3_RAT	P12847 rattus norv
24	129.5	5.1	1427	1	REST_HUMAN	P30622 homo sapien
25	129.5	5.1	2077	1	TEGU_HSV6U	P52340 human herpe
26	129	5.1	539	1	MY53_HYDAT	P39922 hydra atten
27	129	5.1	1853	1	MY5A_MOUSE	Q93104 mus musculu
28	129	5.1	1966	1	MY5B_CAEEL	P02566 caenorhabdi
29	129	5.1	2704	1	BRAL_HUMAN	Q03001 homo sapien
30	128	5.1	1189	1	SCII_CHICK	Q90988 gallus gall
31	128	5.1	1939	1	MYH1_HUMAN	P12882 homo sapien
32	128	5.1	1969	1	MY5A_CAEEL	P12844 caenorhabdi
33	127.5	5.0	1978	1	MYHB_CHICK	P10587 gallus gall

RESULT 1  
CLUS\_HUMAN

ID CLUS\_HUMAN STANDARD; PRT; 449 AA.

AC P10909; P11380; P11381;

DT 01-JUL-1989 (Rel. 11, Created)

DT 01-JUL-1989 (Rel. 11, Last sequence update)

DT 01-MAR-2002 (Rel. 41, Last annotation update)

DE Clusterin precursor (Complement-associated protein SP-40,40)

DE (Complement cytotoxicity inhibitor) (CLI) (NAL and NA2) (Apolipoprotein

DE J) (Apo-J) (TRPM-2).

GN CLU.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

OX NCBI\_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.

RC TISSUE=Liver;

RX MEDLINE=89251601; PubMed=2721499;

RA Kirszbaum L., Sharpe J.A., Murphy B., D'Apice J.F.A., Classon B.,

RA Hudson P., Walker I.D.;

RT "Molecular cloning and characterization of the novel, human

complement-associated protein, SP-40,40: a link between the

complement and reproductive systems.";

EMBO J. 8:711-718(1989).

RN [2]

RP SEQUENCE OF 2-449 FROM N.A.

RX MEDLINE=89386692; PubMed=2780565;

RA Jenne D.E., Tschopp J.;

RT "Molecular structure and functional characterization of a human

complement cytotoxicity inhibitor found in blood and seminal plasma:

identity to sulfated glycoprotein 2, a constituent of rat testis

fluid.";

Proc. Natl. Acad. Sci. U.S.A. 86:7123-7127(1989).

RN [3]

RP SEQUENCE FROM N.A.

RX MEDLINE=90344779; PubMed=1974459;

RA de Silva H.V., Harmony J.A.K., Stuart W.D., Gil C.M., Robbins J.;

RT "Apolipoprotein J: structure and tissue distribution.";

Biochemistry 29:5380-5389(1990).

RN [4]

RP SEQUENCE FROM N.A.

RX MEDLINE=94237156; PubMed=8181474;

RA Wong P., Taillefer D., Lakins J., Pineault J., Chader G.,

RA Tenniswood M.;

RT "Molecular characterization of human TRPM-2/clusterin, a gene

associated with sperm maturation, apoptosis and neurodegeneration.";

Eur. J. Biochem. 221:917-925(1994).

RN [5]

RP SEQUENCE FROM N.A.

RX MEDLINE=93186813; PubMed=7680346;

RA Pineault J.M., Tenniswood M.;

RT "Genomic organization and expression of the rat TRPM-2 (clusterin)

gene, a gene implicated in apoptosis.";

RT J. Biol. Chem. 268:5021-5031(1993).

RN [6]

Q07970 arabidopsis  
Q02440 gallus gall  
P09kx2 homo sapien  
P35579 homo sapien  
P02549 homo sapien  
P49454 homo sapien  
O67124 aquifex aeo  
P26386 saccharomyc  
P38871 saccharomyc  
O51246 borrelia bu  
P29616 gallus gall  
Q9Y411 homo sapien

34 127 5.0 793 1 KATA\_ARATH  
35 127 5.0 1829 1 MY5A\_CHICK  
36 126 5.0 1941 1 MYH2\_HUMAN  
37 126 5.0 1960 1 MYH9\_HUMAN  
38 126 5.0 2418 1 SPCA\_HUMAN  
39 125.5 5.0 3210 1 CENF\_HUMAN  
40 125 4.9 978 1 RA50\_AQUAE  
41 125 4.9 1790 1 US01\_YEAST  
42 124.5 4.9 571 1 YHY4\_YEAST  
43 124.5 4.9 971 1 Y228\_BORBU  
44 124.5 4.9 1102 1 MYSC\_CHICK  
45 124.5 4.9 1855 1 MY5A\_HUMAN

#### ALIGNMENTS

RP SEQUENCE FROM N.A.  
RA Blechschmidt K., Rosenthal A.;  
RL Submitted (Oct-2000) to the EMBL/GenBank/DBJ databases.  
RN [7]  
RP SEQUENCE OF 61-449 FROM N.A.  
RX MEDLINE=92020896; PubMed=1924317;  
RA Danik M., Chabot J.G., Mercier C., Benabid A.L., Chauvin C.,  
RA Ouliron R., Suh M.;  
RT "Human gliomas and epileptic foci express high levels of a mRNA  
RT related to rat testicular sulfated glycoprotein 2, a purported marker  
RT of cell death.";  
RL Proc. Natl. Acad. Sci. U.S.A. 88:8577-8581(1991).  
RN [8]  
RP SEQUENCE OF 140-449 FROM N.A.  
RC TISSUE=Liver;  
RA Glew M.D., Kirsbaum L., Bozas S.E., Walker I.D.;  
RL Submitted (JAN-1993) to the EMBL/GenBank/DBJ databases.  
RN [9]  
RP SEQUENCE OF 23-33 AND 228-240.  
RX MEDLINE=91265608; PubMed=3154963;  
RA Hochstrasser A.-C., James R.W., Martin B.M., Harrington M.,  
RA Hochstrasser D.F., Pometta D., Merrill C.R.;  
RT "HDL particle associated proteins in plasma and cerebrospinal fluid:  
RT identification and partial sequencing.";  
RL Appl. Theor. Electrophor. 1:73-76(1988).  
RN [10]  
RP SEQUENCE OF 23-33; 229-242; 303-317 AND 397-403.  
RX MEDLINE=91230083; PubMed=1903064;  
RA James R.W., Hochstrasser A.-C., Borghini I., Martin B.M., Pometta D.,  
RA Hochstrasser D.F.;  
RT "Characterization of a human high density lipoprotein-associated  
RT protein, NAL/NA2. Identity with Sp-40,40, an inhibitor of complement-  
RT mediated cytotoxicity.";  
RL Arterioscler. Thromb. 11:645-652(1991).  
RN [11]  
RP SEQUENCE OF 23-52 AND 228-257.  
RX MEDLINE=90354412; PubMed=2387851;  
RA de Silva H., Stuart W.D., Park Y.B., Mao S.J.T., Gil C.M.,  
RA Wetterau J.R., Busch S.J., Harmony J.A.K.;  
RT "Purification and characterization of apolipoprotein J.";  
RL J. Biol. Chem. 265:14292-14297(1990).  
RN [12]  
RP SEQUENCE OF 23-37 AND 228-242.  
RX MEDLINE=90097955; PubMed=2601725;  
RA Choi N.H., Mazda T., Tomita M.;  
RT "A serum protein Sp40,40 modulates the formation of membrane attack  
RT complex of complement on erythrocytes.";  
RL Mol. Immunol. 26:835-840(1989).  
RN [13]  
RP SEQUENCE OF 23-41 AND 228-246.  
RX MEDLINE=93319521; PubMed=8328966;  
RA Ghiso J., Matsubara E., Koudinov A., Choi-Miura N.-H., Tomita M.,  
RA Wisniewski T., Frangione B.;  
RT "The cerebrospinal-fluid soluble form of Alzheimer's amyloid beta is  
RT complexed to Sp-40,40 (apolipoprotein J), an inhibitor of the  
RT complement membrane-attack complex.";  
RL Biochem. J. 293:27-30(1993).  
RN [14]  
RP PARTIAL SEQUENCE, DISULFIDE BONDS, AND CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=92201397; PubMed=1551440;  
RA Kirsbaum L., Bozas S.E., Walker I.D.;  
RT "Sp-40,40, a protein involved in the control of the complement  
RT pathway, possesses a unique array of disulphide bridges.";  
RL FEBS Lett. 297:70-76(1992).  
RN [15]  
RP DISULFIDE BONDS.  
RX MEDLINE=93147076; PubMed=1491011;  
RA Choi-Miura N.H., Takahashi Y., Nakano Y., Tobe T., Tomita M.;  
RT "Identification of the disulfide bonds in human plasma protein  
RT Sp-40,40 (apolipoprotein-J).";  
RL J. Biochem. 112:557-561(1992).  
RN [16]  
RP CARBOHYDRATE-LINKAGE SITES.  
RX MEDLINE=97477902; PubMed=9336835;  
RA Kapron J.T., Hilliard G.M., Lakins J.N., Tenniswood M.P., West K.A.,  
RA Carr S.A., Crabb J.W.;  
RT "Identification and characterization of glycosylation sites in human  
RT serum clusterin.";  
RL Protein Sci. 6:2120-2133(1997).  
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
CC CELL DEATH (APOPTOSIS).  
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
DR EMBL; X14723; CAA32847.1; -;  
DR EMBL; M25915; AAA35692.1; ALT\_INIT.  
DR EMBL; J02908; AAA51765.1; ALT\_INIT.  
DR EMBL; M74816; AAA60321.1; -;  
DR EMBL; M64722; AAB06508.1; -;  
DR EMBL; L00974; AAA60567.1; -;  
DR EMBL; AF111103; -; NOT\_ANNOTATED\_CDS.  
DR EMBL; M63379; AAB06507.1; -;  
DR EMBL; M63376; AAB06507.1; JOINED.  
DR EMBL; M63377; AAB06507.1; JOINED.  
DR EMBL; M63378; AAB06507.1; JOINED.  
DR EMBL; A21577; CAA01560.1; -;  
DR PIR; A35833; A35833.  
DR PIR; A37816; A37816.  
DR PIR; B37816; B37816.  
DR PIR; A41386; A41386.  
DR PIR; PLO135; PLO135.  
DR PIR; PLO136; PLO136.  
DR PIR; S04662; S04662.  
DR PIR; S07433; S07433.  
DR PIR; S09339; S09339.  
DR PIR; S34056; S34056.  
DR PIR; S43646; S43646.  
DR SWISS-2DPAGE; P10909; HUMAN.  
DR InterPro; IPR000753; Clusterin.  
DR Pfam; PF01093; Clusterin; 1.  
DR SMART; SM00035; CLA; 1.  
DR SMART; SM00030; CLB; 1.  
DR PROSITE; PS00492; CLUSTERIN\_1; 1.  
DR PROSITE; PS00493; CLUSTERIN\_2; 1.  
KW Apoptosis; Complement pathway; Glycoprotein; Plasma; Sulfation;  
KW Signal.  
FT SIGNAL. 1 22  
FT CHAIN 23 449 CLUSTERIN.  
FT CHAIN 23 227 BETA-CHAIN.  
FT CHAIN 228 449 ALPHA-CHAIN.  
FT CHAIN 102 313 INTERCHAIN.  
FT DISULFID 113 305 INTERCHAIN.  
FT DISULFID 116 302 INTERCHAIN.  
FT DISULFID 121 295 INTERCHAIN.  
FT DISULFID 129 285 INTERCHAIN.  
FT CARBOHYD 86 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 103 103 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 145 145 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 291 291 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 354 354 N-LINKED (GLCNAC. . .).  
FT CARBOHYD 374 374 N-LINKED (GLCNAC. . .).  
FT CONFLICT 28 D -> S (IN REF. 9 AND 10).  
FT CONFLICT 47 Q -> H (IN REF. 11).  
FT CONFLICT 52 G -> Q (IN REF. 11).  
FT CONFLICT 305 C -> M (IN REF. 10).

```
Query Match 17.2%; Score 434; DB 1; Length 449;
Best Local Similarity 25.8%; Pred. No. 1.8e-19;
Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLVIFVCLLWLDKSHCAPTWK-----DKTAISENLKSFSEGEIDAEVKKALT 65
DB 1 MKMTLLFLVGLLL-----TWESQVLGDQTVSDNELQMSNGSKYVKNQIAVAVN 51
QY 66 GIKOMKIMMERKEHTNLMTLKKCKEAKLLNEVQHEERLCRESLADSWG 125
DB 52 GVKQIKTILIEKTEERTLLSNLEAKKEDALNETRESETKLKGPGVCNEMMALWE 111
QY 126 ECRSCLENNCMRIYT-TCQPSWSWSKNIERFFKRIYQFLPFPHEDNEKDLPISEKILIEE 184
DB 112 ECKPCLQKTCMKFYARVCRSGSLVGRLEEFNLQSSPFYFMNGDR-----IDSLLEN 165
QY 185 DAQLTQMEDV----FSQLTVDVNSLNRNFRNFRQMOQEDFQSFHIS-DTDLTPEYF 239
DB 166 DROQTHMLDVMQDHFSSRASSIIDLFODRF-----FTREPQDTY--HYLPFSLPHRRPHF 218
QY 240 FPAFSKEPTMKADLE-OCWDIPNPFQFCNFVSIVSVSETITKMLKATEDLPKQDKAP 298
DB 219 F--FPKSRIVRSLMPPSPYEPLEPHAMFQPFLEMIHEA-----QQAMDIHFSHPAF 267
QY 299 DHGLISKMLPGQDRLGCGELDONLSRCFKHEKCKQCAHLSDC-----PDVPAHLHTEL 354
DB 268 QHPPTPIREGDDRTVCREIRHNSTCGLRMKQCDKCREILSVDCSTNNPSQAKLRREL 327
QY 355 DEAIRLVNSVNOQYQGLQWTRKHELTAYLVKMRGFCGWSELANOAPETELIFNSIQ 414
DB 328 DESQVAERUTRYKNEULLSYQWMLNLSLLEQLNEQFNQVNSRLANLTQGEQYVLRVT 387
QY 415 VVPRHBNISKODETMMTDLISLPSSNFTLKIPLESSESSNFVGVVAKALQHF-KEH 473
DB 388 TVAS-HTSDSDVSGVTEVVVKLFDSQPIVTVVPEVSRKNPRFMEVAKALQHYRKKH 446

RESULT 2
ID CLUS_COTJA STANDARD; PRT; 451 AA.
AC P14018;
DT 01-JAN-1990 (Rel. 13, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DE 15-DEC-1998 (Rel. 37, Last annotation update)
DE Clusterin precursor (51.5 kDa protein).
GN T64.
OS Coturnix coturnix japonica (Japanese quail).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;
OC Coturnix.
OX NCBI_TaxID=93934;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Neuroretina;
RX MEDLINE=89339492; PubMed=2541393;
RA Michel D., Gillet G., Volovitch M., Pessac B., Calothy G., Brun G.;
RT "Expression of a novel gene encoding a 51.5 kD precursor protein is
RT induced by different retroviral oncogenes in quail neuroretinal
RT cells."
RL Oncogene Res. 4:127-136(1989).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=95262670; PubMed=7744033;
RA Michel D., Chatelain G., Herault Y., Brun G.;
RT "The expression of the avian clusterin gene can be driven by two
RT alternative promoters with distinct regulatory elements."
RL Eur. J. Biochem. 229:215-223(1995).
CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY
CC OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,
CC AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED
CC CELL DEATH.
CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER
```

QY 435 LSILPSSNFTLKIPLESSESSNFYGVVAKALQHFKEH 473  
 DB 409 VOLDFEPLSLTPGDISWDDPRFMEIIVAEALQHYKON 447

RESULT 3  
 CLUS\_RABIT STANDARD; PRT; 447 AA.  
 AC Q9XSC5;  
 RC STRAIN-JAPANESE WHITE;  
 DT 16-OCT-2001 (Rel. 40, Created)  
 DT 16-OCT-2001 (Rel. 40, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Clusterin precursor (Apolipoprotein J) (Apo-J).  
 GN CLU.

OS Oryctolagus cuniculus (Rabbit).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.  
 OX NCBI\_TaxID=9986;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC STRAIN-JAPANESE WHITE;  
 RA Miyata M., Biro S., Kaleda H., Eto H., Orihara K., Kihara T.,  
 RA Obata H., Matsushita N., Matsuyama T., Tei C.;  
 RT "Apolipoprotein J/clusterin is induced in vascular smooth muscle cells  
 by balloon injury.";  
 RL Submitted (JAN1999) to the EMBL/GenBank/DBJ databases.  
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
 OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
 AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
 CELL DEATH (BY SIMILARITY).  
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
 CC -1- SUBCELLULAR LOCATION: PRESENT IN CHROMAFFIN GRANULES.  
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL: AF118852; AAD24461.1; -  
 DR InterPro: IPR000753; Clusterin.  
 DR Pfam: PF01093; Clusterin; 1.  
 DR SMART: SM00035; Clb; 1.  
 DR SMART: SM00030; Clb; 1.  
 DR PROSITE: PS00492; CLUSTERIN\_1; 1.  
 KW Glycoprotein; Plasma; Signal.  
 FT SIGNAL 1 22 BY SIMILARITY.  
 FT CHAIN 23 447 CLUSTERIN.  
 FT CHAIN 23 225 BETA-CHAIN (A CHAIN) (BY SIMILARITY).  
 FT CHAIN 226 447 ALPHA-CHAIN (B CHAIN) (BY SIMILARITY).  
 FT DISULFID 101 311 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 112 303 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 115 300 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 120 293 INTERCHAIN (BY SIMILARITY).  
 FT DISULFID 128 283 INTERCHAIN (BY SIMILARITY).  
 FT CARBOHYD 85 85 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 102 102 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 144 144 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 289 289 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 326 326 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 352 352 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 FT CARBOHYD 372 372 N-LINKED (GLCNAC. . .) (POTENTIAL).  
 SQ SEQUENCE 447 AA; 51850 MW; COCA338EEB08590A CRC64;

Query Match 16.4%; Score 413.5; DB 1; Length 447;  
 Best Local Similarity 24.1%; Pred. No. 3e-18;  
 Matches 111; Conservative 93; Mismatches 204; Indels 53; Gaps 11;

QY 34 TWKDKTAISENLKSFSEVGEIDA-----DEEVKALTGKQKIMMERKEHTNLMST 87

Db 13 SWRGQVLGDQLVSDNELQEMSTQSGKYIDREIQNAVGVQEIKTLEKTEWEKRTLLSV 72  
 QY 88 LKCKREEKQEAALKLLNEVQEHLEBEERLCRESLADSWGECRSCLENNCMRIYT-TCOPSW 146  
 Db 73 LEEAKKNKEDALNETROSETKLKAPPEVCNETMMAWEECKPCLKQTCMKFYARVCRSGS 132  
 QY 147 SSVKNKIERFRKIYQFLPFPHEDNEKDLPISEKLIBEDAQLTQMEDVFSOLTVDVNSLF 206  
 Db 133 GLVGRQLEEFNLQSSPFYFWINGDR-----IDSLLENDROQSHVLDV-----MODSF 179  
 QY 207 NRSNVNFMQOQEPDQTFQSHFISDTLTETEPYFFPAESKEPMTKADLFCQWDIP----- 260  
 Db 180 NRATGI---MDLFDQDFRFTHKPQDTFYHSPFSY--FRPPPLHYAKSLVRNIMPLSLYG 234  
 QY 261 --NFFQLCFNFSVSIYESVETITKMLKAIEDLPKQD---KAPDHGGLISKMLPGQDRL 315  
 Db 235 PLNFQDMFQPFEMTHQAQQAQMDVHLHSPAYQTPNVEITGGPD-----DRAV 282  
 QY 316 CGELDNLSRCFKFHEKCKQCOAHLSEDC-----PDVPALHTELDLRLVNVNQYQGI 371  
 Db 283 CKETRHNSGTCLRMKDQCAKQCEILSVDCSANNPSQNLQRLNDLSRLABELTKRYNEL 342  
 QY 372 LQMRKHLEDYALVEKMRGQGVWSELANOAPETETIFNSIQVYVPRITHEGNSKQDETM 431  
 Db 343 LQSYQMKMLNTSSLLDQPNQFNWVSQLANTQGPQDYILRVSTVTS--HTSESEAPSRVT 401  
 QY 432 MTDLSILPSSNFTLKIPLESSESSNFYGVVAKALQHFKE 472  
 Db 402 EVVYKLFDSDDITITPEEVSNDPKNFETVAEKALQEVYK 442

RESULT 4  
 CLUS\_CANFA STANDARD; PRT; 445 AA.  
 ID CLUS\_CANFA  
 AC P25473;  
 DT 01-MAY-1992 (Rel. 22, Created)  
 DT 01-MAY-1992 (Rel. 22, Last sequence update)  
 DT 01-OCT-1996 (Rel. 34, Last annotation update)  
 DE Clusterin precursor (Glycoprotein 80) (GP80).  
 GN CLU.  
 OS Canis familiaris (Dog).  
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.  
 OX NCBI\_TaxID=9615;  
 RN [1]  
 RP SEQUENCE FROM N.A.  
 RC MEDLINE=91236776; Pubmed=2033078;  
 RA Hartmann K., Rauch J., Urban J., Parczyk K., Diel P., Pillarsky C.,  
 RA Appel D., Haase W., Mann K., Weller A., Koch-Brandt C.;  
 RT "Molecular cloning of gp 80, a glycoprotein complex secreted by  
 kidney cells in vitro and in vivo. A link to the reproductive system  
 and to the complement cascade.";  
 RL J. Biol. Chem. 266:9924-9931(1991).  
 CC -1- FUNCTION: NOT YET CLEAR, IT IS KNOWN TO BE EXPRESSED IN A VARIETY  
 OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS, MEMBRANES,  
 AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED WITH PROGRAMMED  
 CELL DEATH.  
 CC -1- SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.  
 CC -1- SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.

-----  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 the European Bioinformatics Institute. There are no restrictions on its  
 use by non-profit institutions as long as its content is in no way  
 modified and this statement is not removed. Usage by and for commercial  
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
 or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
 -----

DR EMBL: M55251; AAA30846.1; -  
 DR PIR: A40018; A40018.  
 DR InterPro: IPR000753; Clusterin.  
 DR Pfam: PF01093; Clusterin; 1.

```
DR SMART: SM00035; Cls: 1.
DR SMART: SM00030; Clb: 1.
DR PROSITE: PS00492; CLUSTERIN_1; 1.
DR PROSITE: PS00493; CLUSTERIN_2; 1.
DR Glycoprotein; Plasma; Signal.
FT SIGNAL 1 22
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT CHAIN 23 445
FT DISULFID 102 309
FT DISULFID 113 301
FT DISULFID 116 298
FT DISULFID 121 291
FT DISULFID 129 281
FT CARBOHYD 86 86
FT CARBOHYD 103 103
FT CARBOHYD 145 145
FT CARBOHYD 277 277
FT CARBOHYD 287 287
FT CARBOHYD 350 350
FT CARBOHYD 370 370
SQ SEQUENCE 445 AA; 51789 MW; 023A37266ABEF374 CRC64;

Query Match 16.2%; Score 410.5; DB 1; Length 445;
Best Local Similarity 23.9%; Pred. No. 4.6e-18;
Matches 117; Conservative 102; Mismatches 192; Indels 79; Gaps 15;

QY 12 MKPPLLVFIVCLWLKSHCAPTKDKTAISENLKSFSEVGEIDAD-----EEVKALT 65
DB 1 MKMTLLLVGLL-----TWNGRVLGQDAVSDTELQEMSTEGSKYINKEIKNALK 51

QY 66 GIKQKIMMERKKEHTNMLSTLKKREKQKQKALKLLNEVQHEERLEERLCRESLADSWG 125
DB 52 GVQIKTLIEQTNEERKSLNLEEAKKKEDALNTKQSETKLKASQGVNCFMFWALWE 111

QY 126 ECRSCLNNCMRYIT-TCQPSWSVNKIRPFERKYYQFLPPHEDNEKDLPISEKLIEE 184
DB 112 ECKPCLKQTCMKFYARVCRSGSLVGHOLEEFLNQSSPFYFWNGDR-----IDSILEN 165

QY 185 DAQLTQMEDVFSOLTVDVNSLFNRNFVFMQOEFDOTFQSHFIS-----DTDLTPEYF 240
DB 166 DROOTHALDV-----MQDSFRASSI-----MDLFDQDFTTREPQDTHYSP--F 209

QY 241 PAFSKEPMTKADLEQCWDI-----PNFQLCFNFSVSIYESVETITKMLKAIE--- 289
DB 210 SLFQRRPFPNPKFRIARNIIPFRFQPLNFHDMFQPFEDMIHQ-A-QQAMDVNLHRIPIYH 268

QY 290 --DLPKQDKAPDHGGLISKMLPCQDRGLCGELDONLSRCFKFHEKCKQKQAHLESDC--- 344
DB 269 PIEFPEED-----NRTVCKEIRHNSGTCLMKMDQCEKQCEILSVDCSSN 312

QY 345 -PDVPALHTELDAIRLVNVSNOQYGOILQMTKRKHLEDYALVEKMRGOGFWGWSLANQA 403
DB 313 NPAQVQLRQELSLSQIAEKFTKLYDELQSQYQEKWFNTSSLKQLKNEQFSWVSQANLT 372

QY 404 PETEIFNSIQVY-PRHEGNTISKQDETMMTDLISLPSSNFTLKIPLIESAESNFIGYV 462
DB 373 QSEDPFYLTQVTSQSDSNVPGVTKV--VKLFDSDPITVMIPEAVSRNNPKFMEIV 430

QY 463 VAKALQHEKE 472
DB 431 AERLQEIYRQ 440

RESULT 5
ID CLUS_MOUSE STANDARD; PRT; 448 AA.
AC Q06890;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Clusterin)
```

```
DE (Apolipoprotein J) (Apo-J).
GN CLJ OR MSGP-2 OR APOJ.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Testis;
RX MEDLINE=93356785; PubMed=8352774;
RA Lee K.-H., Ji Y.-M., Lim H.M., Lee S.-C., You K.-H.;
RT "Molecular cloning and sequencing of sulfated glycoprotein-2 cDNA
RL from testis of mouse: implications of two different mRNAs of SGP-2.";
RL Biochem. Biophys. Res. Commun. 194:1175-1180(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=93359508; PubMed=8354695;
RA French L.E., Chonn A., Ducrest D., Baumann B., Belin D., Wohlwend A.,
RA Kiss J.Z., Sappino A.P., Tschopp J., Schifferli J.A.;
RT "Murine clusterin: molecular cloning and mRNA localization of a gene
RT associated with epithelial differentiation processes during
RT embryogenesis.";
RL J. Cell Biol. 122:1119-1130(1993).
RN [3]
RP SEQUENCE FROM N.A.
RA Hodgdon B.A., Min B.H., Yan H., Farris J.A., Foster D.N.,
RA Strauch A.R.;
RL Submitted (XX-1992) to the EMBL/GenBank/DBJ databases.
RN [4]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6 X CBA; TISSUE=Liver;
RX MEDLINE=94223204; PubMed=8169523;
RA Jordan-Starck T.C., Lund S.D., Witte D.P., Aronow B.J., Ley C.A.,
RA Stuart W.D., Swerfeger D.K., Clayton L.R., Sells S.F., Paigen B.;
RT "Mouse apolipoprotein J: characterization of a gene implicated in
RT atherosclerosis.";
RL J. Lipid Res. 35:194-210(1994).
CC - FUNCTION: NOT YET CLEAR. IT IS KNOWN TO BE EXPRESSED IN A
CC VARIETY OF TISSUES AND IT SEEMS TO BE ABLE TO BIND TO CELLS,
CC MEMBRANES, AND HYDROPHOBIC PROTEINS. IT HAS BEEN ASSOCIATED
CC WITH PROGRAMMED CELL DEATH.
CC - SUBUNIT: ANTIPARALLEL DISULFIDE-LINKED HETERODIMER.
CC - TISSUE SPECIFICITY: MOST ABUNDANT IN STOMACH, LIVER, BRAIN, AND
CC TESTIS, WITH INTERMEDIATE LEVELS IN HEART, OVARY, AND KIDNEY.
CC - PTM: EXTENSIVELY GLYCOSYLATED WITH SULFATED N-LINKED
CC CARBOHYDRATES
CC - SIMILARITY: BELONGS TO THE CLUSTERIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: D14077; BAA03162.1; -
CC EMBL: L08235; AAA37422.1; -
CC EMBL: L05670; AAA37284.1; -
CC EMBL: S70244; AAB30623.1; -
CC PIR: A40714; A40714.
CC PIR: JN0699; JN0699.
CC MGI: 88423; Clu.
CC InterPro: IPR000753; Clusterin.
CC Pfam: PF01093; Clusterin; 1.
CC SMART: SM00035; Cls; 1.
CC SMART: SM00030; Clb; 1.
CC PROSITE: PS00492; CLUSTERIN_1; 1.
CC PROSITE: PS00493; CLUSTERIN_2; 1.
KW Glycoprotein; Sulfation; Signal.
FT SIGNAL 1 21 BY SIMILARITY.
FT CHAIN 22 448 CLUSTERIN.
FT CHAIN 22 226 BETA-CHAIN (SMALL SUBUNIT)
```





FT	CARBOHYD	145	145	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	290	290	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	316	316	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	353	353	N-LINKED (GLCNAC. . .) (POTENTIAL) .
FT	CARBOHYD	373	373	N-LINKED (GLCNAC. . .) (POTENTIAL) .
SEQ	SEQUENCE	446 AA;	51774 MW;	BID5B434B6683AA CRC64;

  

Query Match		15.8%;	Score 400.5;	DB 1;	Length 446;
Best Local Similarity		24.4%;	Pred. No. 1.9e-17;		
Matches	118;	Conservative	98;	Mismatches	196;
				Indels	71;
				Gaps	

  

Qy	16	LLVIFVCLLWLKSHCAPTWK-----DKTAISENLKSFSEYGEIDADEVVKALTGK 168
Db	4	LLLVGLLL-----TWENGPVLGDKRAISDKELQEMSTEGSYVNAKEIKNALKEVK 54
Qy	69	QMKIMMERKEKEHNLMTSLTKKREEKOEAKLLNVEQHLREERLCRESLADSGECR 128
Db	55	QIKTLIEQSNEERKSLSSLEEAKKKEDALNTRTEYKLGSGQLCNETMMALWEECK 114
Qy	129	SCLENNCMRIYT-TCQPSWSVKNKIERFFRKIQYLFPPHEDNEKDLPISEKLIBEDAQ 187
Db	115	PCLKQTCMKFYARVCRSGSLVGHQLEEFNLQSSPFYFWINGOR-----IDSLMENDRQ 168
Qy	188	LTQMEDVFSQLTVDVNSLFNSRVNQMOEQEDQTFQSHFIS---DITDLTEPY----- 238
Db	169	QSHVMDI-----MEDSFNRASNI-----MDELFQRFNREPFDTQFFSGSSH 214
Qy	239	---FFPAFSKEPMTKADLEQCWDIPNFQFLCFNFSYIVSESVETITKMLKAIEDLPKQD 295
Db	215	GSLLFPNPKSFARNIMFPFLFTDL-NVHDMFQPF---FDMIHQAQQAAMDAHLRIPYH- 268
Qy	296	KAPDHGLISKMLP--GQDRGLCGELDONLSRCFKFEKCKQCOAHLSEDC-----PDVPA 349
Db	269	-FPEAG-----VPENSNDRAVCKEIRHNSGTGLRMKDQCEKREITLSVDCSASNSSQM 321
Qy	350	LHTELDIAIRLVNYSNOYGOILLQMTKRLHLEDTAYILVEKMRGFGWVSELNAQAPETEI 409
Db	322	LRQELYSLSLOAEKFKLYDQLLSQYQKMLNFTSLKQLNEQFSWVSQLANTQNDTRY 381
Qy	410	FNSIQVPRTHGENISKQDRTMTDLSILPSSNFTLKIPLESAESSNFTGYVAKALQH 469
Db	382	YLQVTTYV-NSHGSDPSVPSGLITKVVVKLFDSYPTILIPQEVS--DPKFMETVAEEALQQ 438
Qy	470	FKE 472
Db	439	YRQ 441

  

RESULT	8
CLUS_RAT	
ID	CLUS_RAT
AC	STANDARD;
DC	PRT;
DT	447 AA.
DT	01-NOV-1988 (Rel. 09, Created)
DT	01-FEB-1994 (Rel. 28, Last sequence update)
DE	Clusterin precursor (Sulfated glycoprotein 2) (SGP-2) (Dimeric acid glycoprotein) (DAG) (Testosterone repressed prostate message-2) (TRPM 2).
GN	GLU.
OS	Rattus norvegicus (Rat).
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
NCBI_TaxID=10116;	
[1]	
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 22-45 AND 227-241.
RX	MEDLINE=88000523; PubMed=3651384;
RA	Collard M.W., Griswold M.D.;
RT	"Biosynthesis and molecular cloning of sulfated glycoprotein 2
RL	secreted by rat Sertoli cells.";
RL	Biochemistry 26:3297-3303(1987).
RP	[2]
RP	SEQUENCE FROM N.A.







```
DR InterPro; IPR001609; myosin_head.
DR Pfam; PF00612; IQ; 2.
DR Pfam; PF00063; myosin_head; 1.
DR Pfam; PF02736; Myosin_N; 1.
DR Pfam; PF01576; Myosin_tail; 1.
DR PRINTS; PR00193; MYOSINHEAVY.
DR ProDom; PD000355; myosin_head; 1.
DR SMART; SM00015; IQ; 1.
DR SMART; SM00242; MYSC; 1.
DR PROSITE; PS00096; IQ; 1.
KW Myosin; Muscle protein; Coiled coil; Thick filament; Actin-binding;
KW Calmodulin-binding; ATP-binding; Methylation; Alkylation;
KW Multigene family.
FT DOMAIN 1 784 MYOSIN HEAD-LIKE.
FT DOMAIN 785 814 IQ.
FT DOMAIN 814 1938 COILED COIL (POTENTIAL).
FT NP_BIND 179 186 ATP (POTENTIAL).
FT DOMAIN 659 681 ACTIN-BINDING (BY SIMILARITY).
FT DOMAIN 761 775 ACTIN-BINDING (BY SIMILARITY).
FT MOD_RES 130 130 METHYLATION (TRI-) (POTENTIAL).
FT MOD_RES 699 699 ALKYLATION (SH-1) (POTENTIAL).
FT MOD_RES 709 709 ALKYLATION (SH-2) (POTENTIAL).
SQ SEQUENCE 1938 AA; 223678 MW; 1F6D006416381CD5 CRC64;

Query Match 5.7%; Score 144.5; DB 1; Length 1938;
Best Local Similarity 20.7%; Pred. No. 0.29;
Matches 100; Conservative 93; Mismatches 206; Indels 83; Gaps 17;

QY 41 ISENLKSFSEVGEIDAEV-KKALTGKIMMERKEHNLMSTLKKCR-----EEK 95
DB 1150 ISELEAS--GATSAIQEMKKREAEFFKMRDRLEATLQHEATAATLKKQADSVAEL 1207

QY 96 QEALKLLNEVQEHLEEBERLCRESLAD--SWGECRSLENNCKRIVTTCOPSSVKNKI 153
DB 1208 GEQIDNLRQVKLEKSEKSELKWEIDMASNIALSKSNISRTCTVEDQFSEIRAKD 1267

QY 154 ERFRKIYQFLPFPHEDNEKDLPISEKLIEDAQLTQMEDVFSQLVVDVNSLFRSNFNV 213
DB 1268 EQQTQLI-----HDLMQKARLOTQNGELSHRVEKESLIISQATKSKQALTQOLEELK 1320

QY 214 ROMQOQEPD-QTFOSHFTS-----DITDLTEPYFFPAFKEPTMKADLQCV-----DIPNFF 263
DB 1321 RQMEETKAKNAHAHALQSSRHDCDLLR-----EQYEEQEAQALQALSKANSEVAQWK 1376

QY 264 QLCNFSVSIYESVETITKMLKAIEDL-PKQKAPDHGGLISK-----MLPGQDRGLGCEL 319
DB 1377 TKYETAQIQTTELEEAKKLAQRLQEAEEKTETANSKASLEKTKQRLQGEVEDLMDRL 1436

QY 320 DONLSRC-----KFEKCKCKQAHLSDCPDVPAHTLDELAIRLVNVS 365
DB 1437 ERSHTACATLDKKQRFNFKVLAEWKOKLDESQAELEAAQKESLSLTEL---FKMRNAYE 1493

QY 366 QYQGI--LQMTKHLEDYAYLVKMRGFGWYSELANQAPET-----ELI 409
DB 1494 EYVDQLETLRRKNLQEE-----ISDLTEQIAETGKNLQAEKTKKLVQGE 1540

QY 410 FNSIQVPRHGENISKQDETWM---TDLISLPSSNFTLKIPLESAESNFIQYVAKA 466
DB 1541 KSDLQVALBEVGSLEHEESKILRVQELSQVSKSELDKRVIEKDEIEQLKRNQRAAEA 1600

QY 467 LQ 468
DB 1601 LQ 1602

RESULT 13
CENE_HUMAN
ID CENE_HUMAN STANDARD; PRT; 2663 AA.
AC Q02224.
DT 01-JUL-1993 (Rel. 26, Created)
DT 01-JUL-1993 (Rel. 26, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
```

```
DE Centromeric protein E (CENP-E protein).
GN CENPE.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93024922; PubMed=1406971;
RA Yen T.J., Li G., Schaar B.T., Szilak I., Cleveland D.W.;
RT "CENP-E is a putative kinetochore motor that accumulates just before
RT mitosis.";
RL Nature 359:536-539(1992).
RN [2]
RP CHARACTERIZATION.
RX MEDLINE=95196755; PubMed=7889940;
RA Thrower D.A., Jordan M.A., Schaar B.T., Yen T.J., Wilson L.;
RT "Mitotic HeLa cells contain a CENP-E-associated minus end-directed
RT microtubule motor.";
RL EMBO J. 14:918-926(1995).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=98437347; PubMed=9763420;
RA Chan G.K.T., Schaar B.T., Yen T.J.;
RT "Characterization of the kinetochore binding domain of CENP-E reveals
RT interactions with the kinetochore proteins CENP-F and HUBB1.";
RL J. Cell Biol. 143:49-63(1998).
CC -!- FUNCTION: MINUS-END DIRECTED MICROTUBULE MOTOR. PROBABLE
CC KINETOCORE MOTOR. ACCUMULATES JUST BEFORE MITOSIS AT THE G2 PHASE
CC OF THE CELL CYCLE. PROBABLY IMPORTANT FOR CHROMOSOME MOVEMENT
CC AND/OR SPINDLE ELONGATION.
CC -!- SUBUNIT: INTERACTS WITH CENP-F AND BUBRI KINASE.
CC -!- SUBCELLULAR LOCATION: ASSOCIATES WITH KINETOCHORES DURING
CC CONGRESSION, RELOCATES TO THE SPINDLE MIDZONE AT ANAPHASE, AND IS
CC QUANTITATIVELY DISCARDED AT THE END OF THE CELL DIVISION.
CC -!- SIMILARITY: BELONGS TO THE KINESIN-LIKE PROTEIN FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Z15005; CAA78727.1;
DR PIR; S28261; S28261.
DR HSP; P17119; 3KAR.
DR MIM; 117143;
DR InterPro; IPR001752; kinesin.
DR Pfam; PF00225; kinesin; 1.
DR PRINTS; PR00380; KINESINHEAVY.
DR SMART; SM00129; KISC; 1.
DR PROSITE; PS00411; KINESIN_MOTOR_DOMAIN1; 1.
DR PROSITE; PS50067; KINESIN_MOTOR_DOMAIN2; 1.
KW Motor protein; Cell division; ATP-binding; Coiled coil; Mitosis;
KW Cell cycle; Centromere.
KW DOMAIN 1 335 KINESIN-MOTOR.
FT DOMAIN 336 2471 COILED COIL (POTENTIAL).
FT DOMAIN 2472 2663 GLOBULAR (POTENTIAL).
FT NP_BIND 86 93 ATP (BY SIMILARITY).
SQ SEQUENCE 2663 AA; 312087 MW; CEFCL3880C8C8B8 CRC64;
```

```
Query Match 5.6%; Score 142.5; DB 1; Length 2663;
Best Local Similarity 21.5%; Pred. No. 0.56;
Matches 111; Conservative 81; Mismatches 191; Indels 133; Gaps 23;

QY 36 KDKTAISENLKSFSE-----VGEIDA-----DEEVKALTGKIMK----- 71
DB 1200 KKKVKLQKQSFTEHRDLRGYIREIATGLQTKKELKIAHLKHEQETIDELRRSVS 1259
QY 72 -----IMMERKEHTNLMSTLKKREEKQAL---KLLNEVQEHLEEBERLCRESLA- 121
```



Db 1551 KEAASUSELSATETELVOKSTSEGLLDGDLDTETISWAKNVKDLDE-----KKRADLNTTIT 1605  
QY 87 TLKCKREERQKALKLINEVQEHLEERLCRESLADSWGECRSLNNCMRWITTCQPSW 146  
Db 1606 -----ESSAALQNLIESEPI-LEERLC--VLNAGWSRVRTWTDWCNTL----- 1647  
QY 147 SSVKNKIERFFKI-----YQFLPFPHEDNEKDLPISEKIE-----EDAQLTQMED 193  
Db 1648 MHQNOLEIFDGNVAHISTWLYQAEALDEIEKKPTSKQEEIVKRLVSELDDANL-QVEN 1706  
QY 194 VFSOLTVDVNSLNFNSFVFRMQQEFDTQF--SHFISDTDLTETEPFFPAPSKPEMTKA 251  
Db 1707 VDDQALILNAGSSREIVLPEKLAELNRFKVSQHIKSAKL-----LIAQEP----- 1755  
QY 252 DLEQCDWIDPFFQFCNFVSIVSVSETITKMLRAI-----EDLPKQDAPDHGGGLISM 307  
Db 1756 -LYQCLVTTTETFTGVPF--SDLEKLENDIENMLKEVEXHLESSDEDEKMDSESAQIEV 1812  
QY 308 LFGQDRGLGELGDLQNLSCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNVSNOQ 367  
Db 1813 LARGEEMLQHPMEDN-----KKEKIRLQLL-----LLHTRYNK-IRAKPIQQRK 1855  
QY 368 YGQIQLQMKHLEDYAYLVEKMRGQGWSELANQAPETETIIFNSIQVVPRIHEGNISQK 427  
Db 1856 MGOLASGIRSSLPTDYLVE-----INKILLCDDVELSLNVPDELNTAIYE-DFSFQ 1906  
QY 428 DETM 431  
Db 1907 EDSL 1910  
  
RESULT 15  
ID RA50.METJA STANDARD; PRT; 1005 AA.  
AC Q58718;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE DNA double-strand break repair rad50 AtPase.  
GN RAD50 OR M31322.  
OS Methanococcus jannaschii.  
OC Archaea; Euryarchaeota; Methanococcales; Methanococcaceae;  
OC Methanococcus.  
OX NCBI\_TaxID=2190;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=JAL-1 / DSM 2661 / ATCC 43067;  
RX MEDLINE=96337999; PubMed=8688087;  
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,  
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,  
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,  
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,  
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,  
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,  
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,  
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.:  
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus  
jannaschii."  
RL Science 273:1058-1073(1996).  
CC -!- FUNCTION: Involved in DNA double-strand break repair (DSBR). The  
CC rad50/mrell complex possesses single-strand endonuclease activity  
CC and ATP-dependent double-strand-specific exonuclease activity.  
CC Rad50 provides an ATP-dependent control of mrell by unwinding  
CC and/or repositioning DNA ends into the mrell active site (By  
CC similarity).  
CC -!- SUBUNIT: Forms a complex with mrell (By similarity).  
CC -!- SIMILARITY: BELONGS TO THE SMC FAMILY. RAD50 SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announcement/>  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----

DR EMBL: U67572; AAB99331.1; -;  
DR TIGR: M31322; -;  
DR InterPro: IPR003439; ABC\_transportr.  
DR InterPro: IPR001687; ATP\_GTP\_A.  
DR InterPro: IPR003405; SMC\_C.  
DR InterPro: IPR003395; SMC\_N.  
DR Pfam: PF02483; SMC\_C; 1.  
DR Pfam: PF02463; SMC\_N; 1.  
KW DNA repair; Hydrolase; ATP-binding; Coiled coil; Complete proteome.  
FT NP\_BIND 32 39 ATP (BY SIMILARITY).  
FT DOMAIN 158 849 COILED COIL (POTENTIAL).  
SQ SEQUENCE 1005 AA; 119387 MW; 9BBB48173E788F3 CRC64;

Query Match 5.6%; Score 141; DB 1; Length 1005;  
Best Local Similarity 19.8%; Pred. No. 0.22;  
Matches 90; Conservative 80; Mismatches 172; Indels 112; Gaps 17;  
  
QY 20 IYCLLWLDKSHCAPTKWKDTAISENLKSPSEVGEIDAEVKKALGIKQMKIMMERKEK 79  
Db 137 IAKFLSLKPSKLETVAKLIGIDFEKCYQKMGEL--VREYKRLRIEIGELNYKENYEK 194  
QY 80 EHTNLMSTLKKCRKQKQKALKLLNEVQEHLEERLCRESLADSWGECRSLNNCMRWIY 139  
Db 195 ELKNMSQLEERKNKLMKLNKLNKIKKFEDEKLFNE----- 233  
QY 140 TTCQPSWSSVKNKIERFRKIYQFLPFPHEDNEKDLPISE---KLIEEDAQLTQMEDVFS 196  
Db 234 -----WENKLLYEKFINKL-----EERKRALEKQELKILEYDL----- 269  
QY 197 QLTVDVNSLNFNSFVFRMQQEFQD-----TFQSHFISDTDLTETEPFFPAPSKPEP 247  
Db 270 NTVVEARETLNRHKDEYKYSLVDEIRKIESRLKLSHYEDYLLTK-----QLE 321  
QY 248 MTKADLEOCWDIPNFFQFCNFVSIVSVSETITKMLRAIED-LPKQDKAPDHGGGLISK 306  
Db 322 ILKGDIEKLKEFIN-----KSKYRDDIDNLDLTLNKKIKDEIERVETIKD---LLEE 369  
QY 307 MIPGQDRGLGELGDLQNLSCFKFHEKQKQAHLSDECDPVPALHTELDEAIRLVNVSNOQ 366  
Db 370 L-----KNLNEEIEKIEKYKICECKEYKEK-----YLELEKAVEYNKLT 412  
QY 367 QYQIQLQMTF---KHLEDYAYLVEKMRGQGWSELANQAPETETIIFNSIQV---PRIH 420  
Db 413 EYITLQEKKSIEKTNINDLETINKL-----LEETKN--IDIESIENSLKEIEKKKVL 464  
QY 421 EGNISKQDETMTDLSILPSSNFTLKIPLEESAE 454  
Db 465 E-NLQKEKIELNKKKGEINSEIKRLAKILDELKE 497

Search completed: July 2, 2002, 11:50:20  
Job time: 379 sec

**THIS PAGE BLANK (USPTO)**

---



GenCore version 4.5  
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: July 2, 2002, 11:43:23 ; Search time 56.55 Seconds  
(without alignments)  
936.909 Million cell updates/sec

Title: US-09-722-544A-4  
Perfect score: 2529  
Sequence: 1 MRTWDYNSGNMKPPLLVFI.....FIGYVAKALQHEKFEKTKW 477

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*  
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*  
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*  
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*  
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*  
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*  
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*  
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*  
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*  
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*  
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*  
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*  
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*  
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*  
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*  
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*  
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*  
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*  
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*  
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*  
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*  
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

# SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2489	98.4	477	20 AAY30786	Protein encoded by
2	2464	97.4	466	20 AAY30793	Immature human HKN
3	2441	96.5	495	20 AAY30785	Protein encoded by
4	2350	92.9	446	20 AAY30792	Mature secreted hum
5	1757.5	69.5	465	20 AAY30791	Bovine HKNG1 ortho
6	1639	64.8	466	20 AAY30787	Protein encoded by
7	1558	61.6	450	20 AAY30788	Guinea pig HKNG1 o
8	1277	50.5	374	20 AAY30789	Guinea pig HKNG1 o
9	1272.5	50.3	373	20 AAY30790	Guinea pig HKNG1 o
10	780.5	30.9	521	20 AAY30794	Amino acid sequenc
11	434	17.2	449	22 AAU28048	Novel human secret

12	432	17.1	448	12	AA11704	Cytolysis Inhibito
13	405.5	16.0	416	22	ABB50285	Apolipoprotein J o
14	400.5	15.8	446	21	AA03441	Porcine clusterin
15	320	12.7	363	22	AAE03764	Human gene 1 encod
16	173.5	6.9	148	22	AAO00607	Human polypeptide
17	160.5	6.3	116	21	AAO3745	Human secreted pro
18	159.5	6.3	139	22	AAE03783	Human gene 1 encod
19	159.5	6.3	139	22	AAE01703	Human gene 4 encod
20	152.5	6.0	247	22	AAE03787	Human gene 1 encod
21	145	5.7	1372	19	AAW56473	Protein with Rho p
22	143	5.7	1931	22	ABB61012	Drosophila melanog
23	142.5	5.6	2633	22	ABG06505	Novel human diagno
24	141.5	5.6	3433	18	AAW22017	Utrrophin. Homo sa
25	141	5.6	944	21	AAV67600	Human adipose tiss
26	141	5.6	1388	19	AAW56475	Human adipose tiss
27	139.5	5.5	2663	22	AAW39097	Human polypeptide
28	139.5	5.5	2688	22	AAW40883	Human polypeptide
29	139	5.5	1374	22	AAW69070	Human male enhance
30	136	5.4	934	22	AAU01768	Human secreted pro
31	136	5.4	2517	21	AAV71159	Human phosphodiast
32	135.5	5.4	1427	12	AA10534	Human 160kD mediat
33	134	5.3	5373	22	AAU14603	Novel bone marrow
34	134	5.3	5447	22	AAU14697	Novel bone marrow
35	130.5	5.2	2482	16	AAW23996	Human mitotin. Ho
36	130.5	5.2	2482	19	AAW23996	Human mitotin amin
37	130	5.1	990	22	AAW78520	Human protein SEQ
38	130	5.1	1788	22	AAW40467	Human polypeptide
39	128.5	5.1	1392	20	AAW06999	Restin protein seq
40	128.5	5.1	3248	17	AAW99795	Kinetochore protei
41	127.5	5.0	1411	17	AAW02258	Nucleolar/endosoma
42	127.5	5.0	2053	22	AAU03501	Human protein kina
43	127	5.0	1047	18	AAW01535	Cellular homologue
44	126.5	5.0	1780	22	AAW38681	Human polypeptide
45	126	5.0	931	22	AAW79504	Human protein SEQ

## ALIGNMENTS

RESULT 1  
AAY30786  
ID AAY30786 standard; Protein; 477 AA.

AC AAY30786;

DT 23-NOV-1999 (first entry)

DE Protein encoded by a human HKNG1 splice variant HKNG1-VI.

KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.

OS Homo sapiens.

PN WO9947535-A1.

PD 23-SEP-1999.

PF 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

PA (MILL-) MILLENNIUM PHARM INC.  
(REGC) UNIV CALIFORNIA.

PI Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AA210751.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
XX Claim 1; Fig 2A-B; 205pp; English.  
XX  
XX The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)  
XX splice variant. HKNG1 is a gene associated with bipolar affective  
XX disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
XX modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
XX activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
XX activity in individuals can then be administered therapeutically to  
XX treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
XX e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
XX early-onset autosomal dominant myopia. The polynucleotides can be used  
XX in gene therapy techniques to treat such disorders. They are also useful  
XX in diagnosis to identify individuals having, or at risk of developing,  
XX HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
XX mutations especially result in the production of a protein with a  
XX different sequence to the human full-length HKNG1 polypeptide or  
XX splice variant sequences, especially the substitution of a lysine for  
XX a glutamic acid at residue 202 or 184. The polynucleotides are also  
XX useful in gene mapping, to produce probes or primers to identify  
XX similar sequences (e.g. mutants or sequences from different species)  
XX and to produce transgenic animals.  
XX  
SQ Sequence 477 AA;

Query Match 98.4%; Score 2489; DB 20; Length 477;  
Best Local Similarity 99.4%; Pred. No. 6.7e-197;  
Matches 473; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

Qy 1 MRTWDYNSGNKPKPLLVIFVCLLWKDSHCAPTWKDKTAISENLKSFSEVEIDADEV 60  
Db 1 mrtwdysngnmkpllvifvcllwkdshcaptwkdktaisenlksfseveidaadev 60  
Qy 61 KKALTGIKOMKIMMERKEHTNLMSTLKKCKREEKQALKLNEVQHELEERLCRESL 120  
Db 61 kkaltgikmqkimmerkehtnImstlkkckreekqealkllnevqheleerlcresl 120  
Qy 121 ADSNGECSCLENMRYTTCQPSWSSVKNKIERFRKIYQFLPFFHEDNEKOLPISEK 180  
Db 121 adswgecrscleennmcrlyttccqpswssvkakierfrkiyqflpfhnednekdipisek 180  
Qy 181 LIEEDAQLTQMEDVFSQLTVDVNSLFRSFMVROMQOEFQDTFQSHRISDITLPEYFF 240  
Db 181 lieedaqltqmedvfsqltvdvnslnfrsfmvromqgefqtfgskfisdtdltepyff 240  
Qy 241 PAFSKEPTKADLEQCWDIPNFFQLFQNFVSIVESVSETITKMLKATEDLPKQDKAPDH 300  
Db 241 pafskeptkadleqcwdipwffqlfcnfsvsivesvsetitkmlkaiedlpkqdkapdh 300  
Qy 301 GGLTSKMLPGDQGLCGELDONLSRCFKFHEKCKQCAHLSEDCPDVPAHLTELDARL 360  
Db 301 gglltskmlpgdqglcgeIdonlsrckfmeKckqcahlseDCpdvpahteldearl 360  
Qy 361 VNVSNOOQGIQLQTRKHLEDTAYLVEKMGQFGVWSLANQAPETEIFNSIOWVPRIH 420  
Db 361 vnvsnooqgqilqtrkhledtaylvekmrgfgvwselanqapeteiifnsiqvprih 420  
Qy 421 EGNISKQDETMMTDLISLPSSNFTLKIPLESASSSNFIQVYVAKALQHFKEHFT 476  
Db 421 egniskqdetmtdlslpssnftlkipleeasessnfiqyvavakalqhfkehft 476

RESULT 2  
ID AAY30793 standard; Protein; 466 AA.  
XX  
AC AAY30793;  
XX

DT 23-NOV-1999 (first entry)  
DE XX Immature human HKNG1 protein form 2.  
XX  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
XX 16-MAR-1999; 99WO-US05606.  
XX  
XX 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
PA (RESC ) UNIV CALIFORNIA.  
XX  
XX Chen H, Freimer NB;  
XX  
XX WPI; 1999-562047/47.  
XX  
XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
PS Claim 2; Fig 17; 205pp; English.  
XX  
XX The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1  
XX is a gene associated with bipolar affective disorder (BAD). HKNG1  
XX polynucleotides are useful to identify compounds modulating HKNG1 gene  
XX expression or HKNG1 polypeptide expression/activity. Compounds inhibiting  
XX or enhancing HKNG1 gene expression or activity in individuals can then  
XX be administered therapeutically to treat HKNG1-mediated disorders,  
XX especially neuropsychiatric disorders e.g. BAD, schizophrenia, or  
XX HKNG1-mediated myopia disorders, such as early-onset autosomal  
XX dominant myopia. The polynucleotides can be used in gene therapy  
XX techniques to treat such disorders. They are also useful in diagnosis  
XX to identify individuals having, or at risk of developing, HKNG1-mediated  
XX disorders due to mutations in the HKNG1 gene. Such mutations especially  
XX result in the production of a protein with a different sequence to  
XX the human full-length HKNG1 polypeptide or splice variant sequences,  
XX especially the substitution of a lysine for a glutamic acid at residue  
XX 202 or 184. The polynucleotides are also useful in gene mapping, to  
XX produce probes or primers to identify similar sequences (e.g. mutants  
XX or sequences from different species) and to produce transgenic  
XX animals.  
XX  
SQ Sequence 466 AA;

Query Match 97.4%; Score 2464; DB 20; Length 466;  
Best Local Similarity 100.0%; Pred. No. 7.4e-195;  
Matches 466; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 12 MKPPLLVPFIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVEIDADEVKALGKQMK 71  
Db 1 mkppllvpfivcllwlkdsHCaptwkdktaisenlksfseveidaadeevkaltgikmqk 60  
Qy 72 IMMERKEHTNLMSTLKKCKREEKQALKLNEVQHELEERLCRESLADSWGCRSCL 131  
Db 61 immerkehtnImstlkkckreekqealkllnevqheleerlcresladsWGcrscl 120  
Qy 132 ENNCMRIYTTQCPSSWSSVKNKIERFRKIYQFLPFFHEDNEKDLPISEKLEEDAQLTQM 191  
Db 121 enncmriyttcqpsswssvknkierfrkiyqflpfhnednekdipisekleedaqltqm 180

QY 192 EDVFSQLTVDVNSLNFNSFVFRQMOQEFQDTFQSHFISDITDTEPYFFFAFSAKEPMTKA 251  
Db 181 edvfsqitvdvnslnfnsfnvfmqgefqtqfshfisdtdltepyffafsaKepmtka 240  
QY 252 DLEQCWDIPNFQFLFCNFSVIYESVETTKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311  
Db 241 dleqcwdipnffqlfcnfsvsiyesvsettkmlkaiedlpkgdkapdhggliskmlpgq 300  
QY 312 DRGLCGELDONLSRCFKFHEKCKQCAHLSEDCPDVPAALHTEDEARLVNVSNOQYGOI 371  
Db 301 drglcgeldnlsrckfhekckqcahlsecdpdvpallhteidearlvnvsnqyggq 360  
QY 372 LQWTRKHLEDYALVEKMRGQFGVWSELANOAPETEIIFNSIQVVPRIHEGNSKQDETM 431  
Db 361 lqmrkhledtaylvekmrgfgvwselanqapeteiifnsiqvvprihegniskqdetm 420  
QY 432 MTDLSILPSSNFTLIKIPLESASSENFIGYVYVAKALQHFKHEHFTW 477  
Db 421 mtdlsilpssnftlikipleesaessnfigyvvvakalqhfkhehftw 466  
RESULT 3  
AAY30785  
ID AAY30785 standard; Protein; 495 AA.  
XX  
AC AAY30785;  
XX  
DT 23-NOV-1999 (first entry)  
XX  
DE Protein encoded by human HKNG1 cDNA.  
XX  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
PF 16-MAR-1999; 99WO-US05606.  
XX  
PR 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.  
PR 28-OCT-1998; 98US-0106056.  
PR 22-JAN-1999; 99US-0236134.  
XX  
PA (MILL-) MILLENNIUM PHARM INC.  
PA (REGC ) UNIV CALIFORNIA.  
XX  
PI Chen H, Freimer NB;  
XX  
WPI: 1999-562047/47.  
DR N-PSDB; AAZ10750.  
XX  
New HKNG1 polynucleotides useful in diagnosis and treatment of  
PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
PT schizophrenia  
XX  
Claim 1; Fig 1A-B; 205pp; English.  
XX  
The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1  
is a gene associated with bipolar affective disorder (BAD). HKNG1  
polynucleotides are useful to identify compounds modulating HKNG1  
expression or HKNG1 polypeptide expression/activity. Compounds inhibiting  
or enhancing HKNG1 gene expression or activity in individuals can then  
be administered therapeutically to treat HKNG1-mediated disorders,  
especially neuropsychiatric disorders e.g. BAD, schizophrenia, or  
HKNG1-mediated myopia disorders, such as early-onset autosomal  
dominant myopia. The polynucleotides can be used in gene therapy  
techniques to treat such disorders. They are also useful in diagnosis  
to identify individuals having, or at risk of developing, HKNG1-mediated

CC disorders due to mutations in the HKNG1 gene. Such mutations especially  
CC result in the production of a protein with a different sequence to  
CC the human full-length HKNG1 polypeptide or splice variant sequences,  
CC especially the substitution of a lysine for a glutamic acid at residue  
CC 202 or 184. The polynucleotides are also useful in gene mapping, to  
CC produce probes or primers to identify similar sequences (e.g. mutants  
CC or sequences from different species) and to produce transgenic  
CC animals.  
XX  
SQ Sequence 495 AA;  
Query Match 96.5%; Score 2441; DB 20; Length 495;  
Best Local Similarity 98.9%; Pred. No. 6.4e-193;  
Matches 465; Conservative 1; Mismatches 4; Indels 0; Gaps 0;  
QY 7 SNSGNMKPPLLVTIVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGIDAEVKKALTG 66  
Db 25 nnsngnmkppllvtfivcllwlkdschaptwkdktaisenlksfsevgideadeevkkaltg 84  
QY 67 IKQMKIMMERKEKEHTNLMSTLKKCREEKQKALKLLNEVOEHLEEEERLCRESLADSWGE 126  
Db 85 ikqmkimmerkekehtnlmstlkkcreekqkalllnevehleeeerlcresladshge 144  
QY 127 CRSCLENNCMRIYTCOPSNSSVKNKITERFRKIYQFLFPFHEDNEKDLPISEKLIIEEDA 186  
Db 145 crsclenncmriytcopswsvknkierfrkiyqflfpfhednekdpiisekllieeda 204  
QY 187 QLTOMEDVFSQLTVDVNSLNFNSFVFRQMOQEFQDTFQSHFISDITDTEPYFFFAFSKE 246  
Db 205 qltqkdvfsqitvdvnslnfnsfnvfrqkgepdqtfqshfisdtdltepyffpafske 264  
QY 247 PMTKADLEQCWDIPNFQFLFCNFSVIYESVETTKMLKAIEDLPKQDKAPDHGGLISK 306  
Db 265 pmtkadlegcwdipnffqlfcnfsvsiyesvsettkmlkaiedlpkgdkapdhgglisk 324  
QY 307 MLPQDQRLGELDONLSRCFKFHEKCKQCAHLSEDCPDVPAALHTEDEARLVNVSNOQ 366  
Db 325 mlpqdqrldgeldnlsrckfhekckqcahlsecdpdvpallhteidearlvnvsnoq 384  
QY 367 QYGOILQMTKRKHLEDYALVEKMRGQFGVWSELANOAPETEIIFNSIQVVPRIHEGNSK 426  
Db 385 qyggilqmtkrkhledtaylvekmrgfgvwselanqapeteiifnsiqvvprihegnisk 444  
QY 427 QDETMTMTDLSILPSSNFTLIKIPLESASSENFIGYVYVAKALQHFKHEHFT 476  
Db 445 qdetmtmtsilpssnftlikipleesaessnfigyvvvakalqhfkhehft 494  
RESULT 4  
AAY30792  
ID AAY30792 standard; Protein; 446 AA.  
XX  
AC AAY30792;  
XX  
DT 23-NOV-1999 (first entry)  
XX  
DE Mature secreted human HKNG1 protein sequence.  
XX  
KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
KW schizophrenia; splice variant.  
XX  
OS Homo sapiens.  
XX  
PN WO9947535-A1.  
XX  
PD 23-SEP-1999.  
XX  
PF 16-MAR-1999; 99WO-US05606.  
XX  
PR 16-MAR-1998; 98US-0078044.  
PR 05-JUN-1998; 98US-0088312.

PR	28-OCT-1998;	98US-0106056.			
PR	22-JAN-1999;	99US-0236134.			
XX					
PA	(MILL-) MILLENNIUM PHARM INC.				
PA	(REGC ) UNIV CALIFORNIA.				
PI	Chen H, Freimer NB;				
XX					
DR	WPI; 1999-562047/47.				
XX					
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of				
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and				
PT	schizophrenia				
XX					
PS	Claim 5; Fig 17; 205pp; English.				
XX					
CC	The present sequence is encoded by HKNG1 (Hong Kong new gene 1). HKNG1				
CC	is a gene associated with bipolar affective disorder (BAD). HKNG1				
CC	polynucleotides are useful to identify compounds modulating HKNG1 gene				
CC	expression or HKNG1 polypeptide expression/activity. Compounds inhibiting				
CC	or enhancing HKNG1 gene expression or activity in individuals can then				
CC	be administered therapeutically to treat HKNG1-mediated disorders,				
CC	especially neuropsychiatric disorders e.g. BAD, schizophrenia, or				
CC	HKNG1-mediated myopia disorders, such as early-onset autosomal				
CC	dominant myopia. The polynucleotides can be used in gene therapy				
CC	techniques to treat such disorders. They are also useful in diagnosis				
CC	to identify individuals having, or at risk of developing, HKNG1-mediated				
CC	disorders due to mutations in the HKNG1 gene. Such mutations especially				
CC	result in the production of a protein with a different sequence to				
CC	the human full-length HKNG1 polypeptide or splice variant sequences,				
CC	especially the substitution of a lysine for a glutamic acid at residue				
CC	202 or 184. The polynucleotides are also useful in gene mapping, to				
CC	produce probes or primers to identify similar sequences (e.g. mutants				
CC	or sequences from different species) and to produce transgenic				
CC	animals.				
XX					
SQ	Sequence 446 AA;				
Query Match		92.9%;	Score 2350;	DB 20;	
Best Local Similarity		100.0%;	Pred. No. 1.8e-185;		
Matches 446;		Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;
QY	32	ATPWKDKTAISENLKSFSEVGIDEADEVKKALTGKQKIMMERKEHEHTNLMSTLKKC	91		
Db	1	aptwkdktaisenlksfsevgideadeevkkaltgkqmkimrkekehtnlmtlkkc	60		
QY	92	REEKQALKLNEVQHEERLEERLCRESLADSWGECRSCLENNCMRIYTCOPSSSVKN	151		
Db	61	reekqalkllnevqheleerlresladswwgescrscleenncmriytcopsswsvkn	120		
QY	152	KIERFRKIYQFLFFPHEDNEKDLPISEKLIIEEDAQLTQMEDVFSQLTVDVNSLFRNSFN	211		
Db	121	kierfrkiyqflffphednekdplisekllieedaqltqmedvfsqldvnslnfrsn	180		
QY	212	VFRQMQEEDQTFQSHFISDTDLTPYFPFAPSKPMTKADLEQCWDPNFQFQLCNFSV	271		
Db	181	virqmqeefdtqfshfisdtdltpyffpafskpmtkadleqcwdpfnffqlcfnsv	240		
QY	272	SIYESVSEITIKMLKAIEDLPQDKAPDHGGITSKMLPGODRGLGCELDONLSRCFKPHE	331		
Db	241	siyevsetitkmlkaiedlpkqdkapdhggilskmlpgdrglgeidqnlsrckfkhe	300		
QY	332	KCKQCOAHLSEDCPDVPALHTEDEAIRLVNVSNOOYGOILOMTRKHLEDYAYLVERKRG	391		
Db	301	kcqkqahlsecdcpvpahltedeaairlvnvsngyqgillqmtkrkhledtaylvekmrg	360		
QY	392	QFCGWSELANQAPETEIFNSIQVPRTRHEGNIHQDETMWTDLSILPSSNFTLKIPILEE	451		
Db	361	qfgwseelanqapeiteifnsiqvprtrhegnisqgdetmtdmtdlsilpssnftlkiplee	420		
QY	452	SAESSNFIGYVAKALQHFKEHFKTW 477			

Db	421	saessnfigyvvakalqhfkehfkwtw 446			
RESULT	5				
AAAY30791					
ID	AAAY30791	standard; Protein; 465 AA.			
XX					
AC	AAAY30791;				
XX					
DT	23-NOV-1999	(first entry)			
XX					
DE	Bovine HKNG1 ortholog splice variant protein.				
XX					
KW	HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;				
KW	neuropsychiatric disorder; early-onset autosomal dominant myopia;				
XX	schizophrenia; splice variant.				
OS	Bos sp.				
XX					
PN	WO9947535-A1.				
XX					
PD	23-SEP-1999.				
XX					
PF	16-MAR-1999;	99WO-US05606.			
XX					
PR	16-MAR-1998;	98US-0078044.			
PR	05-JUN-1998;	98US-0088312.			
PR	28-OCT-1998;	98US-0106056.			
PR	22-JAN-1999;	99US-0236134.			
XX					
PA	(MILL-) MILLENNIUM PHARM INC.				
PA	(REGC ) UNIV CALIFORNIA.				
XX					
PI	Chen H, Freimer NB;				
XX					
DR	WPI; 1999-562047/47.				
DR	N-PSDB; AAZ10759, AAZ10760, AAZ10761.				
XX					
PT	New HKNG1 polynucleotides useful in diagnosis and treatment of				
PT	neuropsychiatric disorders, e.g. bipolar affective disorders and				
XX	schizophrenia				
PS	Claim 1; Fig 11A-B; 205pp; English.				
XX					
CC	The present sequence is encoded by bovine HKNG1 ortholog splice				
CC	variant. HKNG1 (Hong Kong new gene 1) is a gene associated				
CC	with bipolar affective disorder (BAD). HKNG1 polynucleotides are				
CC	useful to identify compounds modulating HKNG1 gene expression or				
CC	HKNG1 polypeptide expression/activity. Compounds inhibiting or				
CC	enhancing HKNG1 gene expression or activity in individuals can				
CC	then be administered therapeutically to treat HKNG1-mediated				
CC	disorders, especially neuropsychiatric disorders e.g. BAD,				
CC	schizophrenia, or HKNG1-mediated myopia disorders, such as				
CC	early-onset autosomal dominant myopia. The polynucleotides can be used				
CC	in gene therapy techniques to treat such disorders. They are also useful				
CC	in diagnosis to identify individuals having, or at risk of developing,				
CC	HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such				
CC	mutations especially result in the production of a protein with a				
CC	different sequence to the human full-length HKNG1 polypeptide or				
CC	splice variant sequences, especially the substitution of a lysine for				
CC	a glutamic acid at residue 202 or 184. The polynucleotides are also				
CC	useful in gene mapping, to produce probes or primers to identify				
CC	similar sequences (e.g. mutants or sequences from different species)				
CC	and to produce transgenic animals.				
XX					
SQ	Sequence 465 AA;				
Query Match		69.5%;	Score 1757.5;	DB 20;	Length 465;
Best Local Similarity		71.0%;	Pred. No. 1.5e-136;		
Matches 331;		Conservative 57;	Mismatches 77;	Indels 1;	Gaps 1;
QY	12	MKPPLLVFTVCLLWLKDSHCAPTWKDKTAISENLKSFSEVGIDEADEVKKALTGKQK 71			

Db 1 mkppllvivlyllqrdccaptgkdrtsiredpkfsgageidvdeevkalligmqmk 60  
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOHEERLCLRESLADSWGECRSL 131  
 Db 61 ilmerreehsklmrltkkreekqalklmevqehleeeerlcqvslmgsddecksl 120  
 QY 132 ENNCRIYTTQPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLEEDAQLTQM 191  
 Db 121 esdcmrfyttcqswwsmkstiervfrkiyqlfpheddekelpvgekteedvqlmqi 180  
 QY 192 EDVFSQLVTDVNSLFRNSFNFRMQOEFDFQFOSHFISDTDLTEPYFFPAPFKEPTKA 251  
 Db 181 envfsqltvdvgnmshfvkmgqgefdaifsgysmtdsnepyffpafskepakka 240  
 QY 252 DLEQCWDIPNFOLFPCNFVSIVSYSETITKMLKATIEDLPKQDKAPDHGLSKMLPGQ 311  
 Db 241 hpmqswdlpsffqfncfslsvgsvatvtemkaiedlskqkdsahggspssttwpvr 300  
 QY 312 DRGLCGELDQNLSCFHFKEKQCQKQAHLSDEDCPDVPAHTELDIAIRLVNVSNQYQGI 371  
 Db 301 grgicgepgqnsseclgharcqkqdylwadcpavpelytkadealelvnlsnqgyav 360  
 QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANOAPETELIENSIOVVPRIHEGNISKQDET 431  
 Db 361 lqmqthhiedtymekmredqfwwtelasqcpgsenifsfikvvpvgvhegnfskqdek 420  
 QY 432 MTDSLILPSSNFTLKIPLESAESSNFYGVVAKALQHFKEHFKTW 477  
 Db 421 i-disilpssnftltipleesaessdfysmlakavqhkehfksw 465

RESULT 6  
 AAY30787  
 ID AAY30787 standard; Protein: 466 AA.  
 XX AAY30787;  
 AC XX  
 CC XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Protein encoded by the guinea pig HKNG1 ortholog gphkng1815.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 XX Cavia cobaya.  
 XX  
 PN W09947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-0505606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 PR 05-JUN-1998; 98US-0088312.  
 PR 28-OCT-1998; 98US-0106056.  
 PR 22-JAN-1999; 99US-0236134.  
 XX  
 PA (MILL-) MILLENNIUM PHARM INC.  
 PA (REGC ) UNIV CALIFORNIA.  
 XX  
 PI Chen H, Freimer NB;  
 XX  
 DR WPI; 1999-562047/47.  
 DR N-PSDB; AAZ10755.  
 XX  
 XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia  
 XX  
 PS Claim 1; Fig 7A-B; 205pp; English.  
 XX

CC The present sequence is encoded by guinea pig HKNG1 ortholog gphkng1815.  
 CC HKNG1 (Hong Kong new gene 1) is a gene associated with bipolar affective  
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
 CC activity in individuals can then be administered therapeutically to  
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 466 AA;

Query Match 64.8%; Score 1639; DB 20; Length 466;  
 Best Local Similarity 66.7%; Pred. No. 8.9e-127;  
 Matches 312; Conservative 63; Mismatches 89; Indels 4; Gaps 3;

QY 12 MKSPLLVFIYCLLWLDKSHCAPTKWKTALISENLKSFSEVGEIDAEVVKALTIQKMK 71  
 Db 1 mkppllvivlyllqrdccaptgkdrtsiredpkfsgageidvdeevkalligmqmk 60  
 QY 72 IMMERKEHTNLMSTLKKREEKQKALKLLNEVOHEERLCLRESLADSWGECRSL 131  
 Db 61 ilmerreehsklmrltkkreekqalklmevqehleeeerlcqvslmgsddecksl 120  
 QY 132 ENNCRIYTTQPSWSSVKNKIERFRKIYQFLPFHEDNEKDLPISEKLEEDAQLTQM 191  
 Db 121 esdcmrfyttcqswwsmkstiervfrkiyqlfpheddekelpvgekteedvqlmqi 180  
 QY 192 EDVFSQLVTDVNSLFRNSFNFRMQOEFDFQFOSHFISDTDLTEPYFFPAPFKEPTKA 251  
 Db 180 envfsqltvdvgnmshfvkmgqgefdaifsgysmtdsnepyffpafskepakka 240  
 QY 252 DLEQCWDIPNFOLFPCNFVSIVSYSETITKMLKATIEDLPKQDKAPDHGLSKMLPGQ 311  
 Db 240 daepswaipnvqlcnlsfsvgsvatvtemkaiedlskqkdsahggspssttwpvr 300  
 QY 312 DRGLCGELDQNLSCFHFKEKQCQKQAHLSDEDCPDVPAHTELDIAIRLVNVSNQYQGI 371  
 Db 300 drgsdglglnsdscvnrkrcqkqdylwadcpavpelytkadealelvnlsnqgyav 359  
 QY 372 LQMTKRKHEDTAYLVKMRGQFGWVSELANOAPETELIENSIOVVPRI--HGNISKQDE 429  
 Db 360 vgmqthhiedtymekmredqfwwtelasqcpgsenifsfikvvpvgvhegnfskqdek 419  
 QY 430 TMMDLSILPSSNFTLKIPLESAESSNFYGVVAKALQHFKEHFKTW 477  
 Db 420 tvvps-sllpssnftlsspleksagnanfidhvvkvqlqhfekhtw 466

RESULT 7  
 AAY30788  
 ID AAY30788 standard; Protein: 450 AA.  
 XX  
 AC AAY30788;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Guinea pig HKNG1 ortholog gphkng1815 splice variant gphkng7b.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.



CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 XX

SQ Sequence 374 AA;

Query Match 50.5%; Score 1277; DB 20; Length 374;  
 Best Local Similarity 54.5%; Pred. No. 4.8e-97;  
 Matches 255; Conservative 44; Mismatches 73; Indels 96; Gaps 3;  
 QY 12 MKPPLLVFVCLLWKLKSHCAPTKWKTATSENKLSFSEGEIDAEVKKALTIKQMK 71  
 Db 1 mkpllpmpvcllwkchcpcaptwkktaisensansfseagidvgevkialigkqm 60  
 QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEOHEERLCRESLADSGECRSL 131  
 Db 61 immerreehsklmtkkckeeqkalkimnevheleeeslcqvsldswdecrac 120  
 QY 132 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIEDAQLTQM 191  
 Db 121 esncmrfdttcqpawsvkn----- 140  
 QY 192 EDVFSQLTVDVNSLNFNSFNRMQQEFDTQFSHFISDTDLTEPYFFPAFKEPMTKA 251  
 Db 141 -----mepayra 147  
 QY 252 DLEQCWDIPNFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311  
 Db 148 daepswaipnvfqlcnlsfsvyqsvseklittratedppkqdknsqggsilkepq 207  
 QY 312 DRGLCGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTELDDEAIRLVNSNQYQGI 371  
 Db 208 drgsdgkigndscvnrkrcqkcgqylsdccpnvpelyrelnealrlvrsnqgdyq 267  
 QY 372 LQMTKRKHELTAYLVKMRGOGFVWSLANQAPETETIIFNSIOVVPRI--HEGNISKODE 429  
 Db 268 vqmtqyhledtllmekmreqfgwsealayspgaedifnpvkmvalsahegnssddqd 327  
 QY 430 TMMTDLISLPSSNFTLKIPLEESAESSNFYGVYVAKALQHFKEHFTW 477  
 Db 328 tvvps-sllpsnftlsspleksagnanfidhvvekvlgfhkfhktw 374

RESULT 9

AA37090  
 ID AAY30790 standard; Protein; 373 AA.

AC AAY30790;

XX 23-NOV-1999 (first entry)

DE Guinea pig HKNG1 ortholog gphkngl815 splice variant gphkng7d.

XX HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.

XX Cavia cobaya.

XX WO947535-A1.

PN 23-SEP-1999.

XX 16-MAR-1999; 99WO-US05606.

PR 16-MAR-1998; 98US-0078044.

PR 05-JUN-1998; 98US-0088312.

PR 28-OCT-1998; 98US-0106056.

PR 22-JAN-1999; 99US-0236134.

XX (MILL-) MILLENNIUM PHARM INC.

PA (REGC ) UNIV CALIFORNIA.

XX Chen H, Freimer NB;

XX WPI; 1999-562047/47.

DR N-PSDB; AA210758.

XX New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia

PS Claim 1; Fig 10A-B; 205pp; English.

XX The present sequence is encoded by a guinea pig HKNG1 ortholog  
 CC gphkngl815 splice variant gphkng7d. HKNG1 (Hong Kong new gene 1)  
 CC is a gene associated with bipolar affective disorder (BAD). HKNG1  
 CC polynucleotides are useful to identify compounds modulating HKNG1  
 CC gene expression or HKNG1 polypeptide expression/activity. Compounds  
 CC inhibiting or enhancing HKNG1 gene expression or activity in  
 CC individuals can then be administered therapeutically to treat  
 CC HKNG1-mediated disorders, especially neuropsychiatric disorders  
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.

XX Sequence 373 AA;

Query Match 50.3%; Score 1272.5; DB 20; Length 373;  
 Best Local Similarity 54.5%; Pred. No. 1.1e-96;  
 Matches 255; Conservative 45; Mismatches 71; Indels 97; Gaps 4;

QY 12 MKPPLLVFVCLLWKLKSHCAPTKWKTATSENKLSFSEGEIDAEVKKALTIKQMK 71  
 Db 1 mkpllpmpvcllwkchcpcaptwkktaisensansfseagidvgevkialigkqm 60  
 QY 72 IMMERKEHTNLMSTLKKCKREEKQKALKLNVEOHEERLCRESLADSGECRSL 131  
 Db 61 immerreehsklmtkkckeeqkalkimnevheleeeslcqvsldswdecrac 120  
 QY 132 ENNCMRIYTCOPSWSSVKNKIERFRKIYQFLFPFHEDNEKDLPISEKLIEDAQLTQM 191  
 Db 121 esncmrfdttcqpawsvkn----- 140  
 QY 192 EDVFSQLTVDVNSLNFNSFNRMQQEFDTQFSHFISDTDLTEPYFFPAFKEPMTKA 251  
 Db 141 -----mepayra 146

QY 252 DLEQCWDIPNFQFCNFSVSIYESVETITKMLKAIEDLPKQDKAPDHGGLISKMLPGQ 311  
 Db 147 daepswaipnvfqlcnlsfsvyqsvseklittratedppkqdknsqggsilkepq 206  
 QY 312 DRGLCGELDONLSRCFKFHEKQKQAHLSDECDPVPALHTELDDEAIRLVNSNQYQGI 371  
 Db 207 drgsdgkigndscvnrkrcqkcgqylsdccpnvpelyrelnealrlvrsnqgdyq 266  
 QY 372 LQMTKRKHELTAYLVKMRGOGFVWSLANQAPETETIIFNSIOVVPRI--HEGNISKODE 429  
 Db 267 vqmtqyhledtllmekmreqfgwsealayspgaedifnpvkmvalsahegnssddqd 326  
 QY 430 TMMTDLISLPSSNFTLKIPLEESAESSNFYGVYVAKALQHFKEHFTW 477  
 Db 327 tvvps-sllpsnftlsspleksagnanfidhvvekvlgfhkfhktw 373

RESULT 10  
 ID AAY30794 standard; Protein; 521 AA.  
 XX  
 AC AAY30794;  
 XX  
 DT 23-NOV-1999 (first entry)  
 XX  
 DE Amino acid sequence of human HKNG1 splice variant HKNG1-delta7.  
 XX  
 KW HKNG1; Hong Kong new gene 1; bipolar affective disorder; BAD;  
 KW neuropsychiatric disorder; early-onset autosomal dominant myopia;  
 KW schizophrenia; splice variant.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO9947535-A1.  
 XX  
 PD 23-SEP-1999.  
 XX  
 PF 16-MAR-1999; 99WO-US05606.  
 XX  
 PR 16-MAR-1998; 98US-0078044.  
 PR 05-JUN-1998; 98US-0088312.  
 PR 28-OCT-1998; 98US-0106056.  
 PR 22-JAN-1999; 99US-0236134.  
 XX  
 XX (MILL-) MILLENNIUM PHARM INC.  
 PA (REGC-) UNIV CALIFORNIA.  
 XX  
 PI Chen H, Fretmer NB;  
 XX  
 WPI: 1999-562047/47.  
 DR N-PSDB; AAZ10762.  
 XX  
 PT New HKNG1 polynucleotides useful in diagnosis and treatment of  
 PT neuropsychiatric disorders, e.g. bipolar affective disorders and  
 PT schizophrenia.  
 XX  
 PS Claim 1; Fig 18A-B; 205pp; English.  
 XX  
 CC The present sequence is encoded by a HKNG1 (Hong Kong new gene 1)  
 CC splice variant. HKNG1 is a gene associated with bipolar affective  
 CC disorder (BAD). HKNG1 polynucleotides are useful to identify compounds  
 CC modulating HKNG1 gene expression or HKNG1 polypeptide expression/  
 CC activity. Compounds inhibiting or enhancing HKNG1 gene expression or  
 CC activity in individuals can then be administered therapeutically to  
 CC treat HKNG1-mediated disorders, especially neuropsychiatric disorders  
 CC e.g. BAD, schizophrenia, or HKNG1-mediated myopia disorders, such as  
 CC early-onset autosomal dominant myopia. The polynucleotides can be used  
 CC in gene therapy techniques to treat such disorders. They are also useful  
 CC in diagnosis to identify individuals having, or at risk of developing,  
 CC HKNG1-mediated disorders due to mutations in the HKNG1 gene. Such  
 CC mutations especially result in the production of a protein with a  
 CC different sequence to the human full-length HKNG1 polypeptide or  
 CC splice variant sequences, especially the substitution of a lysine for  
 CC a glutamic acid at residue 202 or 184. The polynucleotides are also  
 CC useful in gene mapping, to produce probes or primers to identify  
 CC similar sequences (e.g. mutants or sequences from different species)  
 CC and to produce transgenic animals.  
 XX  
 SQ Sequence 521 AA;  
 XX  
 Query Match 30.9%; Score 780.5; DB 20; Length 521;  
 Best Local Similarity 73.9%; Pred No. 7, 1e-56;  
 Matches 156; Conservative 8; Mismatches 16; Indels 31; Gaps 3;  
 QY 7 SMSGNKKPPLLVFIVCLLMKSHCAPTWKDKTAISENLKSFSEVGEIDAEVKKALFG 66  
 Db :|||||  
 118 nmsgnmkpllvfivcllvklkshcaptwkdktaisenlkfsfsevgeldadeevkkaltg 177

QY 67 IKQKIMMERKEKEHTNLMSTLKKCREEKQALKLLNEVQEHLEEEERLCRESLADSWGE 126  
 Db :|||||  
 178 ikqmkimmerkekehtnlmstlkkcreekqalkllnevqehleeeerlcresladswe 237  
 QY 127 CRSCLENNCMRIYTCQPSWSSVKNKI--ERFRKIIYQFLFPFHEDNEKDLPISEKLEE 184  
 Db :|||||  
 238 crsclenncmriytcqpswssvknklltfeafrcy----- 274  
 QY 185 DAQLTQMEDVFSQLT---VDVNSLNFNSFN 212  
 Db :|||  
 275 ---lgrtdevgnltricqdvsnfmknakv 302  
 RESULT 11  
 AAU28048  
 ID AAU28048 standard; Protein; 449 AA.  
 XX  
 AC AAU28048;  
 XX  
 DT 18-DEC-2001 (first entry)  
 XX  
 DE Novel human secretory protein, Seq ID No 217.  
 XX  
 KW Human; secreted protein; arthritis; Crohn's disease; sepsis; shock;  
 KW ischaemia-reperfusion injury; haematopoiesis; cancer; neuropathy;  
 KW transgenic animal; Alzheimer's disease; Parkinson's disease; burn;  
 KW amyotrophic lateral sclerosis; platelet disorder; thrombocytopenia;  
 KW ulcer; osteoporosis; bone degenerative disorder; periodontal disease;  
 KW gut protection; lung; liver fibrosis; immune deficiency; infection;  
 KW severe combined immunodeficiency; SCID; autoimmune disorder; allergy;  
 KW multiple sclerosis; rheumatoid arthritis; diabetes mellitus; asthma;  
 KW fertility; analgesic; pain; antigen.  
 XX  
 OS Homo sapiens.  
 XX  
 PN WO200166689-A2.  
 XX  
 PD 13-SEP-2001.  
 XX  
 PF 05-MAR-2001; 2001WO-US04942.  
 XX  
 PR 07-MAR-2000; 2000US-0519705.  
 PR 19-MAY-2000; 2000US-0574454.  
 PR 17-JUN-2000; 2000US-0596193.  
 PR 14-JUL-2000; 2000US-0616847.  
 PR 19-SEP-2000; 2000US-0665363.  
 PR 20-OCT-2000; 2000US-0693267.  
 XX  
 PA (HYSE-) HYSEQ INC.  
 XX  
 PI Tang YT, Liu C, Asundi V, Xu C, Wehrman T, Ren F, Ma Y, Zhou P;  
 PI Zhao QA, Yang Y, Drmanac RT, Zhang J, Chen R, Xue AJ, Wang J;  
 XX  
 WPI: 2001-589934/66.  
 DR N-PSDB; AAS44948.  
 XX  
 PT Novel polypeptides and nucleic acids obtained from cDNA libraries  
 PT prepared from various human tissues, for diagnosis and treatment of  
 PT cancer, neurological, inflammatory, and autoimmune disorders -  
 XX  
 PS Example 3; SEQ ID No 217; 107pp; English.  
 XX  
 CC The invention relates to novel isolated human secreted polypeptides (I)  
 CC and polynucleotides (II). (I) and (II) are useful for treating  
 CC inflammatory conditions such as arthritis, nephritis, Crohn's disease,  
 CC ischaemia-reperfusion injury, shock, sepsis, immune responses, and is  
 CC involved in increasing haematopoiesis, stem cell survival, bone growth  
 CC and remodeling. (I), (II) and modulators of (II) are useful for  
 CC prophylaxis or treatment of one or more cancers. (II) is also useful for  
 CC creating transgenic animals useful for studying the in vivo activities of  
 CC the polypeptide as well as for studying modulators of the polypeptides.  
 CC (I) induces the proliferation of neural cells and regeneration of nerve  
 CC and brain tissue and is useful for the treatment of central and



peripheral nervous system diseases and neuropathies, such as Alzheimer's, Parkinson's disease, Huntington's disease, and amyotrophic lateral sclerosis. In addition, (I) is involved in chemotactic or chemokinetic activity, regulation of haematopoiesis and is useful for treating myeloid or lymphoid cell disorders, platelet disorders such as thrombocytopenia and for regeneration of bone, cartilage, tendon, ligament and/or nerve tissue growth, and in tissue repair, healing of burns, incisions, ulcers, for treating osteoporosis, osteoarthritis, bone degenerative disorders, or periodontal disease. Furthermore, (I) is also useful for gut protection or regeneration and treatment of lung or liver fibrosis, reperfusion injury in various tissues, various immune deficiencies and disorders including severe combined immunodeficiency (SCID), bacterial or fungal infections, autoimmune disorders e.g. multiple sclerosis, rheumatoid arthritis, diabetes mellitus, myasthenia gravis, allergic reactions and conditions, such as asthma or other respiratory problems. In addition, (I) affects biorhythms or circadian cycles of rhythms, fertility, metabolism, catabolism, anabolism, storage or elimination of dietary fat, lipid, protein, carbohydrate, vitamins, minerals, provides analgesic effects or other pain reducing effects, immunoglobulin like activity and can act as an antigen in a vaccine composition to raise an immune response. AAU28020-AAU28395 represent novel human secreted protein amino acid sequences of the invention.

Sequence 449 AA;

Query Match 17.2%; Score 434; DB 22; Length 449;  
 Best Local Similarity 25.8%; Pred. No. 2.3e-27;  
 Matches 124; Conservative 96; Mismatches 208; Indels 52; Gaps 14;

QY 12 MKPPLLVIVCLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDADEEKKALT 65  
 Db 1 mmktllilfvgl-----twesqvlqdtvsdnelqemsgskynkeiqnavn 51

QY 66 GIKOMKIMMERKEHTNLMTLKKCKREEKQKALKLLNEVQHEERLCRESLADSWG 125  
 Db 52 gvqkikltktneerktllsneeakkkedalnetresetkikelpgvcnetmalwe 111

QY 126 ECRSCLNNCMRIYT-TCQPSWSVKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLEE 184  
 Db 112 eckpcklqtkmkyarvcrgsglvgqlleeflnqspfyfwmngdr-----idsllen 165

QY 185 DAOLTMEDV----FSQLTVDVNSLFRNFVFRQWQOEFDQTFQSHFIS-DTDLTEPYF 239  
 Db 166 drqqtthmldvmqdhfrassilidelqdrf-----ftrepqdy--hylpfslphrrphf 218

QY 240 FPAFSKEPTKADLE-QCWDIPNFFQFCNFSVSVESVETITKMLKALEDLPKQDKAP 298  
 Db 219 f--fksrivrslmpfspyepflnfhamfqpflenihea-----qqamdihfhsfaf 267

QY 299 DHGGLSKMLPGQDGLGELDONLSRCFKFHEKCKQKQAHLSDEC----PDVPALHTEL 354  
 Db 268 qhpptefiregddrtvcvairhstgclrmkqcdkcreilsvdcstnnpqaklrrel 327

QY 355 DEAIRLVNSNOQYQGLQMTKRLHEDTAVLVEKMGQFCGWSELANOAPETIIENSTQ 414  
 Db 328 desiqvaerltrkynellksyqwmkntssilleqlneqfnwvrlanltqgedqyyrvrt 387

QY 415 VVPRIHGNIISKODEMTMDLSILPSNSFTLKIPLESSESSNFIQYVAKALQHF-KEH 473  
 Db 388 tvas-htsdsvpgsvgtvkvifdsdpitvtvpevrknkpnkmetvaekalqeyrkkh 446

RESULT 12  
 AAR11704  
 ID AAR11704 standard; Protein; 448 AA.  
 AC AAR11704;  
 XX AAR11704;  
 XX AAR11704;  
 DT 20-JUN-1991 (first entry)  
 XX Cytolysis Inhibitor.  
 XX

cytolysis inhibitor; perforin; immunological effector molecule; infertility.  
 Homo sapiens.  
 Key Location/Qualifiers  
 Peptide 1..21  
 Protein 22..226 /label= signal peptide  
 Protein 227..448 /label= A-chain  
 Protein 227..448 /label= B-chain  
 DE39333850-A.  
 18-APR-1991.  
 06-OCT-1989; 89DE-3933850.  
 06-OCT-1989; 89DE-3933850.  
 (SCHD ) SCHERING AG.  
 Tschoopp J, Jenne D;  
 WPI; 1991-118338/17.  
 DNA sequence coding for cytolysis inhibitor - is strong inhibitor of terminal complement protein, eg perforin secreted by killer cells  
 Claim 13; Page 9; 15pp; German.  
 This cytolysis inhibitor is encoded by a 1.7kb BamHI-KpnI fragment isolated from a liver-specific cDNA library. It is a blood plasma component that inhibits immunological effector molecules. It is used for systemic or local treatment of inflammatory or autoimmune diseases mediated by complement or killer cells. It can also be used for detoxification of membrane-active and cytolytic proteins released by bacteria, fungi and insect venoms. The protein has a further use in the treatment of infertility caused by deficiency of cytolysis inhibitor. Monoclonal antibodies directed against the cytolysis inhibitor are also covered by the invention. They are used to isolate or quantify the natural protein in human plasma.  
 See also AAQ11501 and AAQ11502.  
 Sequence 448 AA;

Query Match 17.1%; Score 432; DB 12; Length 448;  
 Best Local Similarity 25.8%; Pred. No. 3.3e-27;  
 Matches 123; Conservative 96; Mismatches 205; Indels 52; Gaps 14;

QY 16 LLVFIVCLWLKDSHCAPTWK-----DKTAISENLKSFSEVGEIDADEEKKALTGIKQ 69  
 Db 4 lllfvgl-----twesqvlqdtvsdnelqemsgskynkeiqnavngvk 54

QY 70 MKIMMERKEHTNLMTLKKCKREEKQKALKLLNEVQHEERLCRESLADSGECS 129  
 Db 55 iktliektneerktllsneeakkkedalnetresetkikelpgvcnetmmalweeekp 114

QY 130 GLENNCMRIYT-TCQPSWSVKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLEEADQL 188  
 Db 115 clkqtkmkyarvcrgsglvgqlleeflnqspfyfwmngdr-----idsllendrqq 168

QY 189 TQMEDV----FSQLTVDVNSLFRNFVFRQWQOEFDQTFQSHFIS-DTDLTEPYFFPAF 243  
 Db 169 thmldvmqdhfrassilidelqdrf-----ftrepqdy--hylpfslphrrphf--f 219

QY 244 SKEPMTKADLE-QCWDIPNFFQFCNFSVSVESVETITKMLKALEDLPKQDKAPDHGG 302  
 Db 220 pksrivrslmpfspyepflnfhamfqpflenihea-----qqamdihfhsfafqhpp 270

QY 303 LISKMLPGQDRGLGCELDONLSRCKFKHEKCKQKQAHLSKEDC-----PDVPALHTLDEAI 358  
Db 271 tefiregddrtvcvrefhnsqclrmkdqcdkcreilsvdcstnpsqaklrreldeai 330  
QY 359 RLNVNSNOQYGOILQWTRKHLEDYALVEKMRGQFGWYSELANOAPETEIFNFIQVPR 418  
Db 331 qvaerltrkynellksyqwkmlntsslleqneqfnwvsvrlanltqggedqylrvttvas 390  
QY 419 IHEGNSKQDETMTDLSLTPSSNFTLKIPLAESSENFYGVVAKALQHF-KEH 473  
Db 391 -htsdsdpsvgvtevvvklfdspitvcvpevsrknkpfmetvaekalqeyrkkh 445

RESULT 13  
ABB50285  
ID ABB50285 standard; Protein; 416 AA.  
XX  
AC ABB50285;  
XX  
DT 08-FEB-2002 (first entry)  
XX  
DE Apolipoprotein J ovarian tumour marker protein, SEQ ID NO:60.  
XX  
KW Ovarian tumour marker gene; human; overexpression; upregulation;  
KW epithelial tumour; cancer; diagnosis; prognosis; disease monitoring;  
KW identification; serous cystadenoma; borderline serous tumour;  
KW serous cystadenocarcinoma; mucinous cystadenocarcinoma;  
KW mucinous cystadenoma; borderline mucinous tumour; endometrioid carcinoma;  
KW undifferentiated carcinoma; clear cell adenocarcinoma; cystadenofibroma;  
KW adenofibroma; Brenner tumour; serial analysis of gene expression; SAGE;  
KW immune response pathway; cell proliferation regulation; protein folding;  
KW membrane localised; secreted; therapeutic target; cytostatic;  
KW gene therapy; vaccine.  
XX  
OS Homo sapiens.  
XX  
PN WO200175177-A2.  
XX  
PD 11-OCT-2001.  
XX  
PF 03-APR-2001; 2001WO-US10947.  
XX  
PR 03-APR-2000; 2000US-194336P.  
XX  
PA (USSH ) US DEPT HEALTH & HUMAN SERVICES.  
XX  
PI Morin PJ, Sherman-Baust CA, Pizer ES, Hough CD;  
DR WPI: 2001-626450/72.  
DR N-PSDB; ABA83111.  
XX  
PT Detecting and identifying ovarian tumor, identifying increased risk for  
PT developing ovarian cancer, and determining effectiveness of ovarian  
PT cancer treatment, by measuring expression level of ovarian tumor marker  
PT gene -  
XX  
PS Claim 23; Page 103-104; 140pp; English.  
XX  
CC The invention relates to methods for diagnosing and prognosing ovarian  
CC tumours in an individual via the detection and measurement of the  
CC expression of ovarian tumour marker genes (ABA83081-ABA83122, ABA83180,  
CC ABA83182 and ABA83184) or segments thereof (ABA83123-ABA83169, ABA83179,  
CC ABA83181 and ABA83183). The methods of the invention are useful for  
CC detecting an ovarian tumour in a patient, for identifying an individual  
CC at increased risk for developing ovarian cancer, in prognostic tests for  
CC assessing the relative severity of ovarian cancer, in tests for  
CC monitoring a patient in remission from ovarian cancer and in tests for  
CC monitoring disease status in a patient being treated for ovarian cancer.  
CC The methods can additionally be used to identify a particular tumour as  
CC being an ovarian tumour (i.e., an epithelial ovarian tumour selected from  
CC serous cystadenoma, borderline serous tumour, serous cystadenocarcinoma,  
CC mucinous cystadenoma, borderline mucinous tumour, mucinous  
CC cystadenocarcinoma, endometrioid carcinoma, undifferentiated carcinoma,

CC clear cell adenocarcinoma, cystadenofibroma, adenofibroma and Brenner  
CC tumour. The ovarian tumour marker genes of the invention were identified  
CC using SAGE (serial analysis of gene expression) and were found to be  
CC overexpressed in a broad variety of ovarian epithelial tumour cells  
CC relative to normal ovarian epithelial cells. The marker genes are  
CC implicated in immune response pathways, in the regulation of cell  
CC proliferation and in protein folding, and many of these are membrane-  
CC localised or secreted. In addition to their use as diagnostic and  
CC prognostic markers, the ovarian tumour marker genes of the invention are  
CC proteins may be used as therapeutic targets for the treatment and  
CC prevention of ovarian cancer. Sequences ABB50257-ABB50299 represent  
CC proteins encoded by ovarian tumour marker genes of the invention.  
XX  
SQ Sequence 416 AA;  
Query Match 16.0%; Score 405.5; DB 22; Length 416;  
Best Local Similarity 25.9%; Pred. No. 4.6e-25;  
Matches 113; Conservative 91; Mismatches 196; Indels 37; Gaps 12;  
QY 49 SEVGIDADEEVKALTGKIKOMKIMMERKEHEHTNLMSTLKKCKREEKQKALKLLNEVQEH 108  
Db 2 snqgskynkqlnavngvkgiktliektneerktllnleakkkkadalnetresetk 61  
QY 109 LEEERLRCRESLADSWGECRSCLENNCMRIYT-TCQPSWSSVKNKIERFFRKIYQFLPFP 167  
Db 62 lkelpgvcnctmmalweeckpolqtkcmkfyarvcrgsglvgvrgleeflnqspfyfwm 121  
QY 168 HDNEKDLPISEKLEEDAQTQMEDV----FSQLTVDVNSLFNRSFNVFRMQQEFQOT 223  
Db 122 ngdr-----idsllendrqqthmldvmqdhfstrasslidelqdrf-----ftrepqdt 170  
QY 224 FOSHEIS-DTDLTEPYFFPAPSKPEWTKADLB-QCWDIPNPFQPCNFSVSIVSETI 281  
Db 171 y--hylpfslphrrphff--fpksrivrslmpfspeplnfhamfqpflmlehea----- 221  
QY 282 TKMLKAIEDLPKQDKAPDGHGLISKMLPGQDRGLGCELDONLSRCKFKHEKCKQKQAHLS 341  
Db 222 ----qgandihfspafghptefiregddrtvcvrefhnsqclrmkdqcdkcreils 277  
QY 342 EDC----PDVPALHTLDEAIRLNVNSNOQYGOILQWTRKHLEDYALVEKMRGQFGWVS 397  
Db 278 vdcstnpsqaklrreldeaiqvaerltrkynellksyqwkmlntsslleqneqfnwvs 337  
QY 398 ELANOAPETEIFNFIQVPRVPRHEGNSKQDETMTDLSLTPSSNFTLKIPLAESSESN 457  
Db 338 rlanitqgedqylrvttvas-htsdsdpsvgvtevvvklfdspitvcvpevsrknkpf 396  
QY 458 FIGYVYVAKALQHF-KEH 473  
Db 397 fmetvaekalqeyrkkh 413

RESULT 14  
AAB03441  
ID AAB03441 standard; Protein; 446 AA.  
XX  
AC AAB03441;  
XX  
DT 03-JAN-2001 (first entry)  
XX  
DE Porcine clusterin protein sequence.  
XX  
KW Pig; clusterin; cell migration; wound healing; angiogenesis;  
KW cancer; vascular trauma; vascular disease; atherosclerosis;  
KW Restenosis; complement cytotoxicity inhibitor; SP-40; 40; apoJ;  
KW testosterone repressed prostate message-2; sulfated glycoprotein-2.  
XX  
OS Sus scrofa.  
XX  
FH Key Location/Qualifiers  
FT Peptide 1..22  
FT /label= signal\_peptide

FT	Protein	23...225	
FT	/label=	mature_clusterin_alpha_chain	
FT	Protein	226..446	
FT	/label=	mature_clusterin_beta_chain	
FT	Region	74..80	
FT	/label=	nuclear_localisation_signal	
FT	Modified-site	86..88	
FT	/label=	potential_glycosylation_site	
FT	Modified-site	103..105	
FT	/label=	potential_glycosylation_site	
FT	Modified-site	145..147	
FT	/label=	potential_glycosylation_site	
FT	Cleavage-site	225..226	
FT	/label=	interchain_cleavage_site	
FT	/note=	"leads to the creation of alpha and beta chains"	
FT	Modified-site	290..292	
FT	/label=	potential_glycosylation_site	
FT	Modified-site	316..318	
FT	/label=	potential_glycosylation_site	
FT	Modified-site	353..355	
FT	/label=	potential_glycosylation_site	
FT	Modified-site	373..375	
FT	/label=	potential_glycosylation_site	
XX	WO200034469-A1.		
PN			
XX			
PD	15-JUN-2000.		
XX			
PF	10-DEC-1999;	99WO-US29262.	
XX			
PR	11-DEC-1998;	98US-0111856.	
XX			
PA	(UUNY ) UNIV NEW YORK STATE RES FOUND.		
XX	Millis AJT;		
PI			
XX			
DR	WPI; 2000-431300/37..		
XX			
PT	Clusterin and gp38K-related peptide capable of altering cell migration		
PT	useful for treating atherosclerosis, cancer and stenosis following		
PT	vascular trauma or disease "		
XX			
PS	Disclosure; Fig 1; 43pp; English.		
XX			
CC	The present sequence is the protein sequence of porcine clusterin.		
CC	Clusterin (also known as complement cytolysis inhibitor, sulfated		
CC	glycoprotein-2, testosterone repressed prostate message-2, SP-40, 40		
CC	ApoJ) is essential for the migration of vascular smooth muscle cells		
CC	(VSMC). The gene and protein can, therefore, be used to promote wound		
CC	healing, angiogenesis and vasculogenesis, in the treatment of stenosis		
CC	following vascular trauma or disease and to treat atherosclerosis, and		
CC	antisenese sequences can be used to treat cancer, as angiogenesis is		
CC	vital for tumour survival.		
XX			
XX	Sequence	446 AA;	
XX			
XX	Query Match	15.8%; Score 400.5; DB 21; Length 446;	
XX	Best Local Similarity	24.4%; Pred. No. 1.3e-24;	
XX	Matches 118; Conservative	98; Mismatches 196; Indels 71; Gaps 1	
QY	16	LLVFIYCLLWKDSHCAPTWK-----DKTATSENLKSFSEVGETDADBEVKKALTGK 68	
Db	4	llllvglll-----twengpwrlgdkaisdkelqemstegskyvkeiknalkevK 54	
QY	69	OKIMMERKEKENTNLMSTLKKREBKQBALKLLNEVQEHLEFEERLCRESLADSGECR 128	
Db	55	qiktliqesneerksllsleeaakkkedalndrtetkklkgsgqlcnetmmalweeck 114	
QY	129	SCLENNCMRIYT--TCQPSWSSVKNKTERFRKTYQFLFFHEDNEKDLPISEKLEEDAQ 187	
Db	115	pclktqcmkfayrcsgsglvgqhgleefngsspfyfwngdr-----idslmndrqr 168	

Search completed: July 2, 2002, 11:43:24  
Job time: 658 sec

Novel human secreted proteins and nucleic acids for diagnosing,  
preventing and treating neurological, cardiovascular, infectious,  
autoimmune, gastrointestinal, bone disorders, cancer, particularly  
ovarian cancer

Claim 11: Page 386-387; 421pp; English.

AA08191-AD08213 represent cDNAs corresponding to 19 human secreted  
protein genes and AA03764-AA03786 represent the proteins they encode.  
AA03787-AA03800 represent human secreted protein fragments or variants.  
The genes and their secreted proteins are useful for preventing,  
treating or ameliorating medical conditions, e.g., by protein or gene  
therapy. Pathological conditions can be diagnosed by determining the  
amount of the new protein in a sample or by determining the presence of  
mutations in the new genes. Specific uses are described for each of the  
19 genes, based on the tissues in which they are most highly expressed,  
and include developing products for the diagnosis or treatment of  
proliferative disorders, cancer, tumours, foetal and developmental  
abnormalities, haematopoietic disorders, diseases of the immune system,  
AIDS, autoimmune diseases (e.g., rheumatoid arthritis), inflammation,  
allergies, neurological disorders (e.g., Alzheimer's disease,  
Parkinson's disease), cognitive disorders, schizophrenia, asthma,  
skin disorders (e.g., psoriasis), sepsis, diabetes, atherosclerosis,  
cardiovascular disorders, angiotensin-related disorders, kidney disorders,  
gastrointestinal disorders, pregnancy-related disorders, endocrine  
disorders, and infections. The proteins can also be used to aid wound  
healing and epithelial cell proliferation, to prevent skin aging due to  
sunburn, to maintain organs before transplantation, for supporting cell  
culture of primary tissues, to regenerate tissues, to identify their  
cognate ligands or binding partners, and in chemotaxis, and can be used  
as a food additive or preservative to modify storage properties.  
Antibodies specific for a protein of the invention can be used in  
alleviating symptoms associated with the disorders mentioned above, and  
in diagnostic immunoassays e.g., radioimmunoassay or enzyme linked  
immunosorbent assay (ELISA). The present sequence represents a human  
secreted protein of the invention.

Sequence 363 AA;

Query Match 12.7%; Score 320; DB 22; Length 363;  
Best Local Similarity 22.6%; Pred. No. 4.3e-18;  
Matches 86; Conservative 80; Mismatches 126; Indels 88; Gaps 12;  
QY 16 LLVFIYCLWLKDSHCAPTWKDTATSENLKSFSEVGEIDAD-----BEVKKALTGIQK 69  
DB 4 LLLLVGLL-----twengrvlgdqmvstdeiqemstegskyinreiknalkgvkq 54  
QY 70 MKTMMERKEHTNLMSTLKKCKEEKQEAALKLLNEVQEHLEEEERLCRESLADSNCEGRS 129  
DB 55 ikllieqtneerkslltnleeeakkkkedalndtkdsemklkasqgvcndtmmlweeckp 114  
QY 130 CLNNNCWRIYT-TCQPSWSVSKNKIERFFRKIYQFLFPFHEDNEKDLPISEKLIEDAQ 188  
DB 115 clkqcmkfyarvcrstglvgqveeflnqspfyfwingdr-----idsllendrqq 168  
QY 189 TQMEDVFSQITVDVNSLFNKNFVFMQOEFDQTQSHFISDTDLTEPYFFPAPFSK--- 245  
DB 169 thaldv-----mqdsfdrassi-----mdelfqdrfft-reaqdpfhfspsfsqr 213  
QY 246 -----EPWTKADLEQCWDIDPFFQFCNFSVSVSVSETITKM 284  
DB 214 rpfffnlkhrfarnlmpfpyqpl-----nfhdmfqpfddmhq-qgandvn 260  
QY 285 LKAIEDLPQDKADPHGLISKMLPQDRLGCLGELQDNLSRCFKFHEKCKQCAHLSEDC 344  
DB 261 lhrhphfpmfteedn-----qdqavckeihrhstgclkmkdqckcreillsdc 310  
QY 345 -----PDVPALHTEDEAIRL 360  
DB 311 ssnpaqvqlrqelnlnslqi 330